THIS WEEK

EDITORIALS

WORLD VIEW What to do with the Russian Academy of Sciences **p.379**

CLIMATE Model warns of a spate of heatwaves to come **p.380**

BONE MEAL Marine worms devour whale skeletons **p.381**

Subject to question

Even when conducting clinical trials to study widely used therapies, researchers must ensure that they disclose the full risks to patients.

Full disclosure of the potential risks to people who volunteer to be test subjects for biomedical research has been a bedrock of ethical protections for decades. Now, a fresh question has come to the fore: how best to protect human subjects in trials that examine the effectiveness of existing therapies that are already in widespread use.

On 28 August, the US office charged with protecting human research subjects will hold an unusual public meeting in Washington DC to tackle this contentious, complex issue, which has polarized the biomedical community in recent months. The Office for Human Research Protections (OHRP), part of the Department of Health and Human Services, is asking for input on how institutional ethics committees — the advisory boards that decide whether proposed trials can go ahead — should assess the risks to people in randomized studies that investigate the risks and benefits of existing treatments for the same condition. Such 'standard of care' trials are likely to become more widespread after being mandated in the 2010 health-care law, so a lot is riding on what the OHRP decides. It might insist that these risks be spelled out on patient-consent forms, even though patients with a particular condition would be taking one or the other medication anyway. Those who argue for looser regulations of such research say that this move could put many volunteers off, because they might mistakenly think that the research itself is adding risk of harm.

The issue has been thrust into the spotlight by a protracted controversy over a study of extremely premature infants, funded by the US National Institutes of Health. From 2005 to 2009, the Surfactant, Positive Pressure, and Oxygenation Randomized Trial (SUPPORT) enrolled 1,316 infants born, on average, 14 weeks early and weighing less than a kilogram. Such infants struggle to breathe because of their immature lungs and so are given extra oxygen from birth. Those in the trial were assigned at random to one of two groups. In one, blood oxygen levels were kept at the higher end of the range used in US hospitals, with the attendant risk of causing an eye disorder called retinopathy of prematurity (ROP) — an abnormal growth of retinal blood vessels that blinds 400–600 US infants every year. In the other group, oxygen levels were kept at the lower end of the range, with the accompanying risks including neurodevelopmental disorders and, some experts in the field believed, death. The goal was to determine the effects of lower or higher oxygen levels on the infants' survival, neurological development and likelihood of developing ROP. In short, the trial sought the sweet spot — the level of oxygen supplementation that would lead to maximum survival without damage.

RISK AVERSE

In 2011 the OHRP, responding to a complaint, began to investigate the informed consent forms signed by parents at the 23 SUPPORT sites. In March this year, it concluded that the forms failed to describe "the reasonably foreseeable risks of blindness, neurological damage and death". All but two of the forms failed to note, for instance, that infants in the group maintained at higher oxygen levels would have a greater chance of eye damage, yet more than half said that infants in the lower-level group could benefit from a lower risk of eye disease or less need for eye surgery. None noted the increased risks of neurodevelopmental disorders in the lower-level group. None listed death as a possible risk of the procedure, although the trial protocol (not seen by parents) did list death among the related adverse events "that may be related to the study". The consent forms did reassure parents that: "Because all of the treatments proposed in this study are within standard of care, there is no predictable increase in risk for your baby."

"Transparency and respect for research subjects must be beyond reproach."

Much of the biomedical establishment has rallied to support the trial investigators and the ethics committees that approved the informed consent forms. They argue that the babies encountered a set of grave risks inherent to being premature, not to being randomly assigned to one or the other arm of the trial. Because the trial administered treatment within accepted guidelines endorsed by the American Academy of Pediatrics, they say, the study added no risk and thus the consent forms were adequate.

The goals of SUPPORT were laudable and addressed a need for better information for physicians. And the study did produce illuminating findings: the infants who received lower levels (aiming to keep the oxygen saturation of their haemoglobin at 85–89%) were less likely to get severe eye disease — but more likely to die — than infants receiving oxygen at 91–95% saturation levels. But in an age in which it is more important than ever that transparency and respect for research subjects must be beyond reproach, the SUPPORT consent forms simply do not pass muster. And although it is true that, collectively, the infants enrolled in the study may have been at no greater risk of a negative outcome than infants who were not enrolled, it is not collectives who sign informed consent documents. It is individuals.

Put yourself in the position of a parent with an extremely premature infant. Would you make the decision to enrol your child in the trial if the consent form stated in simple language that babies assigned to one group were more likely to go blind, and that those in the other were at a higher risk of getting neurodevelopmental disabilities? Equally, would you decide to enrol if the form spelled out that, if you do not take part, your own physician and institution might keep your infant in the middle of the range, trying to avoid either outcome? Perhaps you might, but you would do so with full knowledge of the attendant risks. The parents in this case could not do so.

In June, under pressure from many sides, the OHRP said that it would not sanction the SUPPORT investigators and instead would hold next week's meeting. No matter the thorniness of the issues raised there, research is still research in whatever context, and the duty to protect human subjects must remain paramount. ■


JULIA PANCHENKO

What is to be done about Russian science?

Government reforms to the Russian Academy of Sciences have met with controversy, but some form of change is needed, argues **Mikhail Gelfand**.

The reforms to the Russian Academy of Sciences that were announced by the Russian government in June were met with almost unanimous opposition in the scientific community. Critics have complained that the severity of the proposed changes — which include transferring properties owned by the academy into the hands of the government — is combined with a vagueness about how they will be implemented. Furthermore, the abrupt announcement came with political pressure and a smear campaign in state-owned media, but without public debate. The government response is that all opinions have been stated already many times (which is partially true), that the reform has only just started and the detail will be clarified later, and that the only way to move forward is, well, to move forward.

The reform bill is currently in the state Duma, where it will receive its final reading next month. The signals from the Duma are inconclusive: although some members, including the speaker, Sergei Naryshkin, mentioned the possibility of returning the bill to the second-reading stage, where substantial amendments are possible, there have been no official statements along these lines.

The government has managed to achieve the seemingly impossible: it has brought Russian science together. Academic stalwarts who oppose any change (aside from an increase in the academy's budget) have united with proponents of (reasonable) reform, long-time critics of the academy and scientists who normally run shy of politics. Despite summer vacations, some members of the scientific community are discussing the post-reform system, and others are planning meetings and strikes that aim to overturn the proposals. A meeting of all groups working on projects that relate to the reforms is scheduled for the end of August.

Some of the ideas being discussed seem more realistic than others. With its head firmly in the sand, the presidium of the academy has prepared a list of amendments to the bill that mainly aim at returning to the pre-reform status quo. Another working group, formed by the Scientific Council of the Ministry of Science (independent researchers who are largely critical of the reforms) and the Society of Scientific Researchers (an independent, informal society with free membership that is restricted only by a publications-based qualification) has offered other suggestions. These tackle fundamental issues such as whether Russian science should be arranged around institutes or laboratories, what the balance should be between guaranteed and grant-based funding, and whether academy research should be subject to international review. At their heart, these discussions debate whether the future of the academy is as a learned society, similar to the UK Royal Society, or as a Soviet-style 'ministry of basic sciences' that manages and funds its institutes.

➲ NATURE.COM
Discuss this article online at:
go.nature.com/scrf8y

One burning problem acknowledged by most of those working on possible alternatives to the government reform is the future working relationships among the academy, its institutes and a new agency set up by the proposed law to handle academy property. The bill provides no details and is unclear about whether this property includes the land, buildings and equipment that are directly used for research purposes. In particular, scientists worry that all purchases will need to be approved by bureaucrats with no understanding of science.

In the words of a famous Russian novel, what is to be done? A prerequisite for successful reform is the creation of a transparent funding system that also features regular international assessment of laboratories, institutes and large projects. A more strategic goal should be to mend the split between research and higher education. This should not be done by simply increasing the financing of universities at the cost of research institutes, but rather by encouraging the educational activity of institute researchers and the research activity of university professors. Specific grant-based support of joint projects between institutes and universities is also needed. The teaching load of university professors, which is currently much higher than that of professors in the West, must be decreased, and regular audits should cover not only the academy but also other research centres. Finally, Russia's leaders need to understand that science cannot be expected to produce immediate results in the form of 'innovations', but instead needs to be judged on its own merits.

THERE IS GENERAL AGREEMENT IN RUSSIA THAT CHANGE IS OVERDUE.

Ultimately, deep reform can be implemented only if the government has a popular mandate for change. This is not the case in Russia. Hence all reforms are met with distrust and a search for a hidden agenda. This distrust has been fuelled by a project to incorporate several physics institutes into the Kurchatov Institute, the head of which, Mikhail Kovalchuk, is widely believed to be a trusted adviser of President Vladimir Putin. The centre enjoys a steady, rich flow of finance, despite a scientific output that is much weaker than academic institutes. The Institute of Theoretical and Experimental Physics, formerly one of the top scientific institutions, all but ceased to function when it was incorporated into the Kurchatov Institute.

There is general agreement in Russia that change is overdue; even the new academy leadership acknowledges this. Forms of change separate from the unpopular government proposals could work. But it is unclear whether research in Russia can make the shift, given the current political climate and the academy's systemic, deeply rooted problems. ■

Mikhail Gelfand *is a professor of the Faculty of Bioengineering and Bioinformatics, Moscow State University at and deputy director of the Institute for Information Transmission Problems of the Russian Academy of Sciences.*
e-mail: mikhail.gelfand@gmail.com

RESEARCH HIGHLIGHTS

Selections from the scientific literature

JULIE DERMANSKY/CORBIS

CLIMATE SCIENCES

Global heat waves on the rise

Heat waves will become more common by 2040. Climate models used by Dim Coumou of the Potsdam Institute for Climate Impact Research in Germany and Alexander Robinson at the Complutense University of Madrid predict that about 20% of Earth's land surface will experience monthly temperatures that are more than three standard deviations from the mean. Such extremes occur over about 5% of the global land surface today, and were seen in the 2012 heat wave across the United States and in the Texan heat wave of 2011, when reservoirs nearly dried up (pictured).

The heat-wave projections stand until 2040, no matter how much more carbon dioxide humans put into the air. After that, lowered emissions could allow temperatures to stabilize, whereas maintaining current emissions would see the frequency of heat waves continue to rise.

Environ. Res. Lett. 8, **034018 (2013)**

ANIMAL BEHAVIOUR

How termites drum up help

To call for assistance when their nest is under attack, some termites use their heads.

Wolfgang Kirchner and Felix Hager of Ruhr University in Bochum, Germany, mimicked predator attacks on two species of African termites that grow fungi in long underground galleries connected to their nests. Specialized soldier termites responded by drumming their heads against the ground, which drew more soldiers to the alarm. Laboratory experiments confirmed that soldiers sense the low-frequency vibrations.

Vibrations from simulated drumming dissipated within 40 centimetres, but many galleries are much longer. To transmit alarm calls over greater distances, the termites pass signals on to others until the messages reach soldiers, the researchers suggest.

J. Exp. Biol. 216, **3249–3256 (2013)**

ECOLOGY

Soil life predicts nutrient flow

Studies of soil organisms are usually lab-based, but in a rare field study, Franciska de Vries — now at Lancaster University, UK — and her colleagues looked at the relationship between soil food webs and carbon and nitrogen entering and leaving controlled areas. The 60 sites — in Sweden, the United Kingdom, the Czech Republic and Greece — included grassland, intensely farmed sites and areas with crop rotation.

Intensive land use, such as wheat cultivation, reduced the mass of soil life of all kinds. But the researchers found that biomass within soil was a better predictor of nutrient cycling and soil health than was land usage, and suggest that nutrient models should pay more attention to what happens underground.

Proc. Natl Acad. Sci. USA **http://dx.doi.org/10.1073/pnas.1305198110 (2013)**

CANCER IMAGING

Chemical reaction reveals tumours

A chemical-imaging technique may one day allow tracking of prostate cancer without the need for invasive biopsies.

A team led by Sarah Nelson at the University of California, San Francisco, exploited differences in how healthy and cancer cells break down certain chemicals, using them to spot tumours in 31 human patients.

The researchers used magnetic resonance imaging to observe isotopically labelled pyruvate, a compound that supplies energy to cells. Shortly after injecting the labelled pyruvate into patients, researchers could observe it being converted into lactate in prostate tumours, and the conversion sometimes revealed cancer in regions that had been overlooked by conventional imaging. Signals that were more intense indicated faster metabolism of pyruvate, a property that has been linked in animal studies to more-aggressive forms of cancer.

Sci. Transl. Med. 5, **198ra108 (2013)**

MATERIALS

Catalyst forms under pressure

High pressure normally turns the porous minerals known as zeolites into a powdery, non-crystalline mess. Chemists have now shown that this is not

always the case, by converting a zeolite into a stable new mineral using high-pressure compression.

Zeolites are often used as catalysts because their pores can trap a range of molecules. Depending on zeolite structure, the minerals can break up heavy oil, separate out gases or purify water.

In the hunt for fresh zeolite structures, Avelino Corma at the Polytechnic University of Valencia, Spain, and his co-workers used diamond anvils to compress the minerals. At 32,000 times atmospheric pressure, a type of silica zeolite transformed irreversibly into another porous structure, which was better at separating propene and propane than its parent form.

Angew. Chem. **http://dx.doi.org/10.1002/anie.201305230 (2013)**

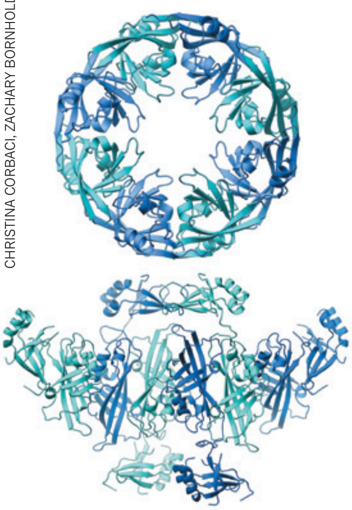
STRUCTURAL BIOLOGY

Lethal viral shape-shifter

An ebolavirus protein adopts drastically different conformations (**pictured**) throughout its life cycle, allowing the deadly virus to do more with fewer genes.

Ebolaviruses kill up to 90% of the people they infect. Erica Ollmann Saphire of the Scripps Research Institute in La Jolla, California, and her colleagues used crystallography, biochemistry and microscopy to track the structure of VP40, a protein that controls ebolavirus assembly and exit from host cells. They learned that the protein does not travel alone as previously thought, but moves to the cell membrane in butterfly-shaped pairs, which then align end-to-end into hexamers that form filaments essential for viral assembly and release. The team also analysed requirements for VP40 to form yet another structure, a previously observed ring that binds to RNA and regulates viral genes in infected cells.

Cell 154, **763–774 (2013)**



CHRISTINA CORBACI, ZACHARY BORNHOLDT, ERICA OLLMANN SAPHIRE/TSRI

CANCER BIOLOGY

Mouth microbe causes cancer

Certain bacteria living in the mouth and gut can invade intestinal cells and trigger changes that lead to colorectal cancer.

A team led by Wendy Garrett at the Harvard School of Public Health in Boston, Massachusetts, found that the bacterium *Fusobacterium nucleatum* induced colonic tumours in genetically susceptible mice.

Separately, Yiping Han at Case Western Reserve University in Cleveland, Ohio, and her colleagues showed that FadA, an adhesion molecule produced by *F. nucleatum*, interacts with a counterpart on mammalian cells and triggers proliferation of colorectal-cancer cells. Colon tissue from patients with tumours had 100 times more copies of the gene encoding FadA than did tissue from healthy individuals.

Cell Host Microbe 14, **195–206; 207–215 (2013)**

PSYCHIATRIC GENETICS

Common variants behind disorders

The risk of getting a psychiatric illness is largely heritable — and many of the genetic variants involved seem to be shared across disorders.

The international Cross-Disorder Group of the Psychiatric Genomics Consortium identified common genetic variants in more than 30,000 patients diagnosed with one of five psychiatric disorders, and compared these with thousands of non-diagnosed controls.

These variants accounted for 17–29% of risk for the illnesses, and there is substantial overlap between disorders. For example, in schizophrenia, 15% of the variants overlapped with bipolar disorder, 9% with depression and 3% with autism.

Nature Genet. **http://dx.doi.org/10.1038/ng.2711 (2013)**

COMMUNITY CHOICE

The most viewed papers in science

NEUROSCIENCE

Commitment beats will

HIGHLY READ on www.cell.com in August

Avoiding temptation is more effective than resisting it.

Molly Crockett, now at University College London, and her colleagues tested 78 men as they relied on willpower (resisting an available temptation) or precommitment (voluntarily restricting access to temptation) to obtain rewards.

After rating a set of erotic images, subjects could choose to view a less-enjoyable image immediately or a more-enjoyable one after a delay. In willpower tasks, the option to see the less-preferred image was always available, whereas in precommitment tasks, men chose at the outset whether to wait for a preferred image. Participants were more likely to gain the superior reward in precommitment scenarios, with the benefits of precommitment varying across individuals. Imaging of a subset of 20 men revealed that different brain areas were used for precommitment and willpower.

Neuron 79, **391–401 (2013)**

ZOOLOGY

Bone-eating worms in icy seas

THOMAS DAHLGREN, ADRIAN GLOVER

Two species of bone-devouring worms have been discovered in the cold waters of the Antarctic. Other members of this genus had previously been found only at warmer latitudes.

Scientists led by Thomas Dahlgren at the company Uni Research in Bergen, Norway, found a new species of worm (*Osedax antarcticus*; **pictured**) carpeting whale bones that the team had placed on the sea floor. Another *Osedax* species was found on bones left in shallower water.

Pine and oak planks placed with the bones remained in near-pristine condition, free of the marine invertebrates that usually feed on wood in warmer waters and quickly consume sunken ships. As a result, the researchers suggest that shipwrecks on the cold sea floor will stay remarkably well-preserved.

Proc. R. Soc. B 280, **20131390 (2013)**

For a longer story on this research, see go.nature.com/kb2kix

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

POLICY

No money for eggs

California governor Jerry Brown vetoed a proposed law on 13 August that would have allowed payments to women who donate their eggs for scientific research — a move that may deter other states from attempting to ease similar bans. The measure would have boosted the availability of human eggs for research in fields such as cloning or somatic-cell nuclear transfer. Separate rules prohibit the California Institute for Regenerative Medicine in San Francisco, the state's stem-cell agency, from funding research on stem-cell lines created with eggs from paid donors. See go.nature.com/xkelfv for more.

Nuclear waste

A US appeals court has ruled that the country's Nuclear Regulatory Commission must revive its review of the Department of Energy's application to open a nuclear-waste repository at Yucca Mountain in Nevada. The energy department had sought to withdraw its application in March 2010, after years of political controversy and concerns over whether the site would leak radioactive waste (see *Nature* **473**, 266–267; 2011). But on 13 August, the court ruled that the regulatory commission must continue reviewing the energy department's application.

Hydroelectric halt

India's Supreme Court has ordered that the construction of additional hydroelectric dams must be suspended in the Himalayan state of Uttarakhand. After flash floods and landslides in June killed thousands of people in the state, environmental groups blamed the recent rapid expansion of hydroelectric projects as a contributing factor. On 13 August, the court ordered that an expert group should be set up to study the environmental impact of dams, tunnels and deforestation associated with hydroelectric plants, and to assess whether such projects precipitated June's tragedy. See go.nature.com/pjvsp4 for more.

Amazon drilling

Ecuador's President Rafael Correa has abandoned an initiative intended to persuade wealthy nations to pay his country not to drill for oil in an Amazon rainforest reserve. Correa had asked for US$3.6 billion — 50% of the estimated revenue from development — in exchange for protection of the Yasuní National Park. After receiving only $13 million in pledges since the initiative was mooted in 2007, Correa announced on 15 August that the government would move forward with drilling.

JUAN RENDON

Carnivore misidentified for decades

A nocturnal, tree-dwelling mammal with a bushy tail and teddy-bear-like face is the first new carnivore species to be identified in the Western Hemisphere in 35 years. Dubbed the olinguito (*Bassaricyon neblina*, infant pictured), the 75-centimetre-long inhabitant of the Andean cloud forests had been mistaken in museums and zoos for a close relative — the olingo — for more than a century. Zoologists who reported the finding on 15 August (K. M. Helgen *et al. ZooKeys* **324**, 1–83; 2013) first discovered the mix-up by studying decades-old museum samples, eventually confirming their findings by tracking a live olinguito in Ecuador in 2006. DNA tests revealed that an olingo kept in US zoos during the 1960s and 1970s was actually an olinguito. See go.nature.com/acqawh for more.

BUSINESS

Bribery probe

In a widening crackdown, China will launch a three-month bribery probe across multiple business sectors, including the pharmaceutical and medical-services industries, the country's State Administration for Industry and Commerce said on 14 August. Earlier this month, the authorities began probing Sanofi, based in Paris, over claims that the company offered 1.69 million renminbi (US$274,000) in bribes to physicians in China. In July, the Chinese government opened an investigation into senior executives of GlaxoSmithKline in China for allegedly bribing officials and physicians in the country to boost drug sales.

India drug patents

Following media reports, Swiss drug-maker Roche has confirmed to *Nature* that it will stop enforcing an Indian patent that would have protected its breast-cancer drug trastuzumab (Herceptin) until 2019. Makers of generic pharmaceuticals may now sell cheaper versions of the drug, easing tensions over

drug prices that have seen India reject some patents and skirt others by granting local 'compulsory licences' (see *Nature* **500,** 266; 2013). By avoiding compulsory licensing, the compromise may help Roche to maintain long-term access to the Indian market.

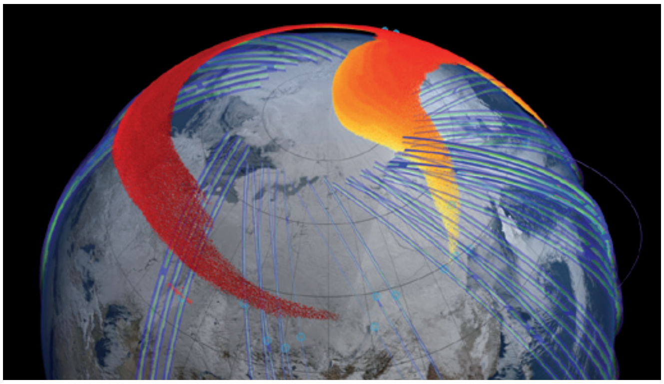
RESEARCH

Planet-naming code

Long-promised guidelines for the public naming of planets and moons were issued by the International Astronomical Union on 14 August. The Paris-based organization, which oversees planetary nomenclature, asks that any group gathering candidate names follow the guidelines. Among other things, the rules discourage using the names of pet animals and forbid the collection of money in the naming process. They were prompted in part by the actions of Uwingu, a space-education company in Boulder, Colorado, that in February asked the public to pay to vote on candidate planet names (see *Nature* **496,** 407; 2013).

Space dust trails

NASA GODDARD'S SCIENTIFIC VISUALIZATION STUDIO



NASA scientists have tracked a dust cloud that was dumped in Earth's stratosphere by a meteor explosion in February over Chelyabinsk in Russia, the agency announced on 14 August. Using satellite data and atmospheric models, researchers found that the space dust — estimated to weigh hundreds of tonnes — reached an altitude of 40 kilometres within hours of the blast. It then swirled around the Northern Hemisphere for days, forming a band (**pictured**) that lasted for at least 3 months. The findings will be published in the journal *Geophysical Research Letters*.

Kepler kaput

NASA is abandoning attempts to resuscitate the crippled Kepler Space Telescope, the agency announced on 15 August. Engineers have spent months trying to repair two of the telescope's four gyroscope-like wheels, which are crucial for controlling its movement. The first wheel failed in July 2012, with a second one breaking in May (see go.nature.com/4w1ufr). The spacecraft needs at least three working wheels to carry out its search for Earth-like exoplanets that might support life. Although Kepler completed its primary mission in 2012, it had begun an extended mission that was scheduled to end in 2016.

PEOPLE

Forensics clash

Senior forensic scientist Wang Xuemei has resigned as vice-president of the Chinese Forensic Medicine Association, it was reported on 18 August. In an online resignation video, she said that she could no longer be associated with the academic organization behind "ridiculous and irresponsible conclusions", referring to her doubts over the association's determination that a man died accidentally in 2010 by falling on to electrified subway tracks in Beijing. She added that she had resolved to quit the forensic system in China. Wang is well known for criticizing last year's conviction of Gu Kailai for poisoning a British businessman, although Wang's video did not refer to this. Gu is the wife of ousted Chinese politician Bo Xilai, whose corruption trial begins this week.

Research fraud

A researcher at Leiden University Medical Center in the Netherlands has been fired for committing scientific fraud, the centre announced on 14 August. Annemie Schuerwegh, who worked in the rheumatology department, admitted manipulating data included in a study published in *Proceedings of the National Academy of Sciences* in 2010, says a report from the centre. She went into the laboratory outside office hours and added mouse antibodies to tubes of human blood samples. The centre will withdraw the article and another paper, and has halted a clinical trial based in part on the fraudulent data.

COMING UP

23 AUGUST
Hoping to bring the International Linear Collider to Japan, the country's high-energy-physics community holds a press conference to announce the site it has chosen to host the proposed atom smasher.

28 AUGUST
The Office for Human Research Protections hosts a public meeting in Washington DC to discuss guidelines on informed consent for clinical trials that test the high and low ranges of standard medical practice (see page 377).
go.nature.com/lb42by

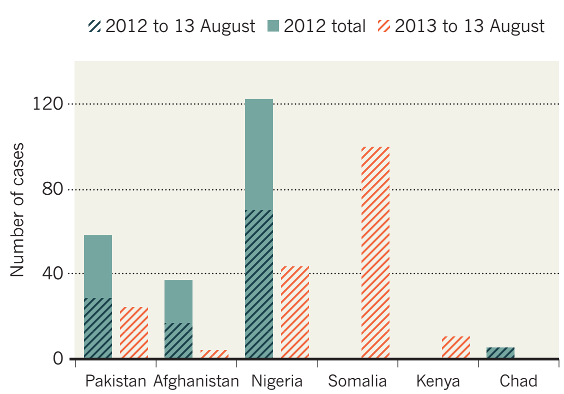
TREND WATCH

SOURCE: WWW.POLIOERADICATION.ORG

Numbers of cases of polio are falling in Nigeria, Pakistan and Afghanistan, countries where wild poliovirus is endemic. But outbreaks have occurred in Somalia and Kenya, part of a band of African countries in which imported poliovirus tends to cause periodic reinfections. In Somalia, emergency vaccinations could prove particularly difficult. The medical charity Médicins Sans Frontières said on 14 August that it would close all its programmes there because of attacks on staff.

POLIO FLARES UP IN SOMALIA

A severe outbreak of polio in the Horn of Africa is worrying public-health officials.



➲ NATURE.COM
For daily news updates see:
www.nature.com/news

NEWS IN FOCUS

PUBLISHING Head online, there are more free papers there than was thought **p.386**

POLICY Romanian science in free fall with hard-won reforms unpicked **p.388**

BIOTECHNOLOGY Gaps in laws could ease path for niche GM crops **p.389**

STEM CELLS Lab-made sperm and eggs spawn hope — and ethical morass **p.392**

STEWART CAIRNS/*NEW YORK TIMES*/REDUX/EYEVINE

Storm runoff can introduce huge amounts of silt into the reservoir system that feeds New York's taps.

WATER MANAGEMENT

Forecasts turn tide on silt

New York pioneers system to protect drinking water from adverse weather events.

BY JEFF TOLLEFSON

When Hurricane Irene hammered the eastern United States in August 2011, floods sent a glut of silt into New York City's drinking-water system. The turbid waters rushed more than 100 kilometres through an aqueduct from the Catskill Mountains to the Kensico Reservoir, the last stop before the slurry would have reached millions of taps. For more than eight months afterwards, the city was forced to use an environmentally contentious chemical to rid the water of silt.

With the frequency and intensity of such siltation events on the rise, New York City is about to embark on a pioneering upgrade to its water system. The focus will not be on new dams or silt traps. Instead, starting this November, New York's reservoirs will be managed by souped-up software that automatically incorporates short-term weather forecasts and seasonal climate forecasts — helping water managers to deal with floods and droughts.

Hydrologists far beyond the Big Apple will be watching closely. New York's programme depends on a streamflow-forecasting system developed by the US National Weather Service, which aims to implement the system nationwide in the coming years. "This project opens the door to a more quantitative use of seasonal climate forecasts, which will help people make better decisions," says Andrew Wood, a hydrologist at the National Center for Atmospheric Research in Boulder, Colorado. Australia is now experimenting with a similar forecasting system, and the European Commission Joint Research Centre has also rolled out a flood-prediction system.

But New York will be one of the first cities to connect the forecasts to a water-management system. Its immediate goal is to manage storm runoff to meet water-quality standards without adding costly new infrastructure to the system, a network of 19 reservoirs and 3 lakes that collectively hold more than 2 trillion litres of water. Under normal circumstances, managers handle siltation by holding water in reservoirs and letting silt settle out. But several times since 2005, the city has had to let silty water flow all the way to Kensico, where it was then treated with aluminium sulphate — a chemical that causes silt particles to coagulate and sink. State and federal regulators have raised concerns about build-up of aluminium sulphate in reservoir sediments, and its potential effects on fish and aquatic organisms.

➲ NATURE.COM
Read more in *Nature*'s special on water resources:
go.nature.com/vrtrdc

The new system is designed to help the city to cope with major siltation episodes, which are expected to increase as the climate warms. With better warning of impending storms, water managers can drain certain reservoirs ahead of time and alleviate the potential for siltation. The software also takes seasonal forecasts into account, allowing managers to work out ▶

various conservation strategies if, for example, drier weather is expected.

"For years, we've done this by ourselves, just trying to balance all of this in our heads," says Jim Porter, who heads water operations for the city's department of environmental protection. "Hopefully, we can predict out a little further into the future."

The potential savings are enormous. Coping with the silt problem by building a new intake system at one reservoir or increasing the size of a second reservoir would cost between US$200 million and $500 million. A new filtration plant could run to more than $10 billion. By contrast, the city's analysis suggests that an integrated reservoir-management system can tackle the problem for roughly $8 million.

But to make it happen, the city first needs better streamflow forecasts. Although the National Weather Service makes streamflow predictions, it has until now done so mainly by comparing current conditions — precipitation, soil moisture, snowpack and streamflow — with historical averages, and then extrapolating the results. This approach assumes that streamflows will evolve as they have in the past under similar circumstances, but does not look ahead to future conditions.

The incoming system — years in the making — combines short-term and seasonal precipitation forecasts, and adds those predictions into the streamflow forecasts. To validate the system, the National Weather Service checked its predictions against historical data.

New York is paying the Weather Service about $1 million to accelerate the process so that the system will be available for use this year. "We now have something that is ready for prime time," says John Schaake, a hydrologist and independent consultant in Baltimore, Maryland, who helped to develop the streamflow-forecast system.

The system will be available at 5 of the 12 regional US river-forecast centres, although it is unclear when it will become standard nationally. There are budget constraints, and each centre will have to customize the system. "There's a lot of interest, but the question is how you institutionalize that," says Kevin Werner, a hydrologist at the Colorado Basin River Forecast Center in Salt Lake City, Utah.

And it is also unclear whether others will follow New York's lead and hitch the forecasts to a reservoir-management system. Demonstrating that the forecasts improve water management should help to ease doubts, says Daniel Sheer, president of HydroLogics Incorporated in Columbia, Maryland, which is providing New York with the reservoir-management software. "There will be much broader interest in the forecasts if we can show that they work." ■

PUBLISHING

Half of 2011 papers now free to read

Boost for advocates of open-access research articles.

BY RICHARD VAN NOORDEN

Search the Internet for any research article published in 2011, and you have a 50–50 chance of downloading it for free. This claim — made in a report[1] produced for the European Commission — suggests that many more research papers are openly available online than was previously thought. The finding, released on 21 August, is heartening news for advocates of open access. But some experts are raising their eyebrows at the high numbers.

There has been a steady move over the past few years towards getting research papers that are funded by government money into the public domain, and the best estimates[2,3] for the proportion of papers free online run at around 30%. But these are underestimates, argues Éric Archambault, the founder and president of Science-Metrix, a consultancy in Montreal, Canada, that conducted the analysis for the European Commission.

The firm initially asked a team led by Stevan Harnad, an open-access campaigner and cognitive scientist at the University of Quebec in Montreal, to check a random sample of 20,000 papers published in 2008 (from the Scopus database of papers run by Elsevier). It used a program designed by Yassine Gargouri, a computer scientist at the same university, to find free articles. The team found that 32% of the papers that it downloaded in December 2012 were freely available. But when Archambault's group checked 500 of these papers manually using Google and other search engines and repositories, the figure rose to 48%.

On the basis of this initial test, Science-Metrix applied its own automated software, or 'harvester', to 320,000 papers downloaded from 2004 to 2011; the tool searches publishers' websites, institutional archives, repositories such as arXiv and PubMed Central, and sites such as the academic networking site ResearchGate and the search engine CiteSeer[X].

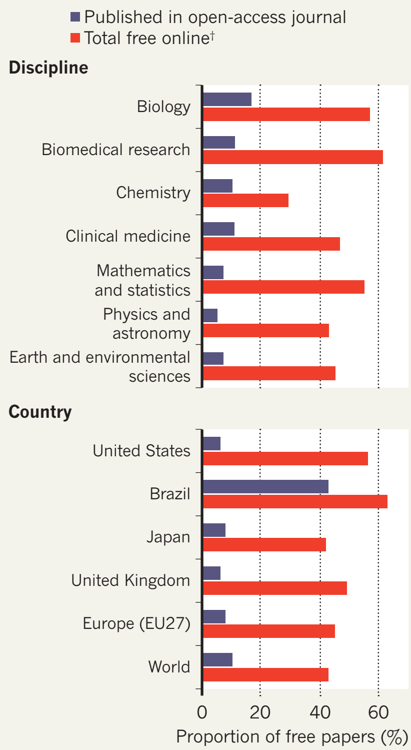
It found that an average of 43% of articles published during 2008–11 are available online for free, with the results varying by country and discipline (see 'Freedom online'). But the true figure is probably higher, because the harvester does not pick up every free paper. When the incompleteness is adjusted for, the proportion of free articles from 2011 rises to about 50%, says Archambault.

The report "confirms my optimism", says Peter Suber, director of the Office for Scholarly Communication at Harvard University in Cambridge, Massachusetts, and a proponent of open access to research. He thinks that it reflects the experiences of working scientists today. "When researchers hit a paywall online, they turn to Google to search for free copies — and, increasingly, they are finding them," he says.

The rise of open-access journals is part of the explanation: the share of papers published in these journals rose from 4% in 2004 to 12% by 2011, the report found — agreeing with figures published last year by Bo-Christer Björk, who

FREEDOM ONLINE

At least* 43% of research papers published during 2008–11 are now free online, but the proportion varies by country and discipline.



*The true proportion is likely to approach 50%, because automated software does not find every free paper.
†Including papers made free by subscription journals, as well as self-archived free papers.

SOURCE: SCIENCE-METRIX

➲ NATURE.COM
For more on open-access publishing, visit: go.nature.com/gdtvaw

LAURIE CAMPBELL/NHPA/PHOTOSHOT

studies information systems at the Hanken School of Economics in Helsinki.

But the number of peer-reviewed manuscripts made free by other means has also increased, the report says. That includes those eventually made free — often a year after publication, and sometimes on a temporary promotional basis — by publishers that charge for subscription. But it also includes manuscripts that researchers themselves archive online on repositories and personal websites. Some of the articles, although free to read, may not meet formal definitions of open access because, for example, they do not include details on whether readers can freely reuse the material.

The report does not try to distinguish between types of manuscript, nor where and how they were posted, says Archambault. "The situation is so complex that it's very hard to measure."

Björk says that the latest measurements seem to have been carefully done, although he adds that because he does not have details of the robotic harvester's code, he cannot evaluate its method. "Experts on the subject would probably agree that the open-access share of papers, measured around a year and a half after publication, is currently at least 30%," he says. "Anything above that is dependent on ways of measuring, with this new study representing the highest estimate."

The report, which was not peer reviewed, calls the 50% figure for 2011 a "tipping point", a rhetorical flourish that Suber is not sure is justified. "The real tipping point is not a number, but whether scientists make open access a habit," he says.

Harnad thinks that the next step should be to obtain more accurate measures of when papers become free. "It's hardly a triumph if articles are only accessible after a one-year embargo," he says. Greater measurement accuracy is tricky to achieve, he adds, because Google routinely blocks all robotic harvesters. He believes that research on the growth of open access should be given special concessions.

The proportion of free online papers is likely to increase in the next few years. The European Commission says that, from 2014, the results of all research funded by the European Union must be open access. And in February, the US White House announced that government-funded research should be made free to read within 12 months of publication (see *Nature* **494,** 414–415; 2013). Federal agencies are due to submit their plans for achieving this to the US Office of Science and Technology Policy by 22 August. ■

1. Archambault, E. *et al. Proportion of Open Access Peer-Reviewed Papers at the European and World Levels — 2004–2011* (Science-Metrix, 2013).
2. Laakso, M. & Björk, B.-C. *BMC Med.* **10,** 124 (2012).
3. Björk, B.-C., Laakso, M., Welling, P. & Paetau, P. *J. Am. Soc. Inf. Sci. Technol.* (in the press).

Soay sheep have greatest sexual fitness when they have two versions of a gene that determines horn size.

EVOLUTIONARY GENETICS

Big horns clash with longevity in sheep

Gene for small horns lowers sexual fitness but boosts lifespan.

BY EWEN CALLAWAY

Alpha Red 78 — a ram with horns like elephant tusks — sired 95 lambs before he died at the ripe (for a ram) old age of nine. A gene with a role in horn growth explains his fertility and his longevity, finds a study of sheep on a remote Scottish isle. The work also explains how variation can persist in traits that offer big reproductive boosts.

Ample horns are a ram's ticket to reproductive success. During the breeding season, males fight for access to females, and those with the largest horns win. But if big horns are a sexual asset, the genes underlying the trait should have become ubiquitous, says Susan Johnston, an evolutionary biologist at the University of Edinburgh, UK, who led the research. Yet some male sheep have short horns or none at all. "From an evolutionary perspective, it doesn't really make sense," Johnston says.

Johnston's team turned to the sheep living on Hirta, an island 160 kilometres west of the Scottish mainland. The animals, a primitive breed called Soay (*Ovis aries*), are known for their diminutive size and their agility on cliffs.

Two years ago, Johnston's group reported that a single gene, *RXFP2*, explains horn variability in the sheep (S. E. Johnston *et al. Mol. Ecol.* **20,** 2555–2566; 2011). One version of the gene, Ho^{+}, is linked to large horns; another allele, Ho^{P}, is associated with small ones.

In the latest study, published in *Nature*, Johnston's team related the *RXFP2* genes of 1,750 sheep to three factors: horn size, reproductive success and lifespan (S. E. Johnston *et al. Nature* http://dx.doi.org/10.1038/nature12489; 2013). Males with one or two copies of the Ho^{+} allele had the biggest horns. They fathered twice as many lambs as those with two copies of the short-horned allele, averaging 3 (versus 1.6) each year, says Johnston. But where lifespan was concerned, rams with two copies of Ho^{P} had an edge, she says, with a 75% chance per year of surviving the harsh Hirta winter, compared with a 61% chance for those with two long-horned alleles.

The scientists found that rams with one version of each allele (heterozygotes) had the best of everything: they were big-horned, fecund and long-lived. And this explains why short-horned rams persist. "I'm just impressed by the simple elegance of this story," says Hopi Hoekstra, an evolutionary geneticist at Harvard University in Cambridge, Massachusetts.

Johnston says that to learn more, scientists will need to study the gene: in humans and mice, it is involved in sexual development and bone density. She adds that heterozygotes such as Alpha Red 78 end up with more offspring largely because they outlive homozygous big-horned males, which tend to die young.

The ram probably wasn't winning on his looks. "He was quite an ugly sheep," she says. ■

RAZVAN CHIRITA/MEDIAFAX PHOTO

Research minister Mihnea Costoiu and his predecessor Ecaterina Andronescu have left scientists unhappy.

POLICY

Romanian science in free fall

Researchers rue the reversal of positive reforms.

BY ALISON ABBOTT

After 11 years away from Romania, developmental biologist Ioan Ovidiu Sirbu thought carefully before returning home to continue his scientific career. He had been convinced that reforms to Romania's cronyism-ridden research landscape were solid, particularly when, in 2011, government grants were for the first time ever allocated solely on the basis of performance.

"With such a fair granting system, I was sure that I could do my research just as well in Romania as in Germany," says Ovidiu Sirbu. "But what happened was really disappointing."

Just months after Ovidiu Sirbu established himself at the Victor Babeş University of Medicine and Pharmacy in Timişoara in 2012, a new government slashed research funding and unpicked the reforms, eliminating rules designed to establish a meritocracy.

Ovidiu Sirbu's disappointment is widely shared. In April, hundreds of scientists took to the streets in protest, and more than 900 signed a petition addressed to Prime Minister Victor Ponta, demanding that the research budget and quality control be restored. The entire National Research Council, Romania's main research-funding agency, resigned in protest (see *Nature* **496,** 274–275; 2013).

With no compromise from the government and the council seats still unfilled, Romanian science is adrift. Scientists are resigned to treading water, in the hope that the tide will turn.

Many of Romania's best researchers left during the political chaos that followed the collapse of communism in 1989. But in 2011, the government passed a law designed to drive up standards in education and science. Research and education minister Daniel Funeriu furnished the law with rules and regulations crafted to break through local power networks and ensure that funding and academic positions would go to the best people — for example by requiring grant applications to be reviewed by foreign experts, and by instituting minimum qualifications for job candidates (see *Nature* http://doi.org/bp7nsg; 2011). At the same time, the research budget was boosted by nearly half.

But that government fell last year. Reversals to the reforms followed; many scientists blame Funeriu's successor, Ecaterina Andronescu.

Andronescu, who took the post last July, is a powerful figure in Romanian academic politics. She was research and education minister in two previous governments, and was rector at the Polytechnic University of Bucharest until she stepped down under Funeriu's conflict-of-interest rules. The law forbade rectors from being politicians, in a bid to stop academics using political positions to help cronies.

Andronescu lost her ministerial post in fresh elections last December, but during her last days in office, she overhauled Funeriu's regulations using three legal tools — including an 'emergency ordinance', or decree, whose rationale, she declared, was a need to make standards attainable to more people. This decree is currently

MISCONDUCT ALLEGATIONS

Plagiarism in politics

Ecaterina Andronescu, Romania's former research and education minister, is a co-author of three out of four papers that international experts claim were plagiarized and infringe copyright. The allegations — the latest in a series of plagiarism scandals to rock Romania (see *Nature* http://doi.org/ngq; 2012) — were published last December on the scientist-run website Integru.org, which was set up with the aim of purging plagiarism from Romanian academia. The editors of the journals — Elsevier's *Thin Solid Films* and *Journal of the European Ceramic Society*, the *IEEE International Semiconductor Conference* and the Romanian Academy's *Romanian Journal of Information Science and Technology* — were told a few days later.

The papers, published in 2006–07, have still not been retracted, says David Tománek, a physicist at Michigan State University in East Lansing, who reads Integru. He told *Nature* that the journals had not replied to requests for information from concerned scientists.

Only one editor — of *Thin Solid Films* — responded to *Nature*'s own requests for information. Joe Greene, a physicist at the University of Illinois at Urbana-Champaign, wrote in an e-mail on 28 May that Elsevier had applied its own investigation procedures. He expected the inquiry to be concluded "within weeks to a couple of months".

Andronescu, who has not responded to *Nature*'s written and telephone requests for comment on this particular issue, denied plagiarism in an interview published in a Romanian newspaper in May, adding that a representative from the *Journal of the European Ceramic Society* had called her just days earlier to congratulate her on a paper. **A.A.**

being discussed in parliament. Andronescu's actions reversed safeguards against conflicts of interest and cronyism. Andronescu did not respond with comment on these specific concerns in time for *Nature*'s press deadline.

Under the latest rules, university rectors can once more be members of parliament, and academics over the retirement age of 65, including Andronescu herself, can hold leadership positions at universities — previously banned to stop people holding on to power for too long. Funeriu had limited academics to supervising eight PhD students at a time — to stop powerful professors from dominating the training of the next generation — but that restriction has now been lifted. And grant applications no longer require review by scientists outside Romania.

In addition, Andronescu abolished the requirement that professors pass a special exam, and loosened Funeriu's minimum criteria for holding an academic post. Critics slam the new criteria as too soft, and say that they are distorted in some subjects — in biology, for example, the focus is on publication of books rather than of peer-reviewed papers. The loosened criteria were applied this year in appointing 1,300 professors as part of Romania's first competition for academic posts since 2009.

Andronescu, who remains a member of parliament, is now head of the senate's education committee and leads her university's senate. She told *Nature* that responsibility for developing minimum criteria for academic appointments lies with the Romanian National Council for the Attestation of University Titles. The criteria then become official through ministerial order.

Late last year, Andronescu was embroiled in further controversy, owing to accusations of plagiarism and copyright infringement in her research papers (see 'Plagiarism in politics').

Her ministerial successor, Mihnea Costoiu, told *Nature* that all procedures for academic appointments had been correctly followed in this year's hiring surge. He added that assertions that standards for becoming a professor had been lowered were "either a misinterpretation or an uninformed assumption on the part of the 'initiators' of this theory".

In April, Costoiu made deep retrospective cuts to the basic research budget, roughly halving the value of grants awarded in 2011 that were already in progress, and stalling the next round of already-evaluated grants. He also launched a new grant competition, for collaborations between research and industry, using laxer rules. Costoiu says that he intends ongoing grants to receive their full monies in later years.

In spite of the tumult, Ovidiu Sirbu remains optimistic. A grant that he applied for in 2012 finally came through this month, although it had been cut by about one-quarter. And he thinks that by staying, he can make a small difference to science in Romania. "One of the good things is that I can train people the way they should be trained," he says. "That's the best I can do now in this country." ■

SKYE HOHMANN BOTANICA/ALAMY

Apple scab is one of several crop diseases that researchers want to beat with genetic engineering.

BIOTECHNOLOGY

US regulation misses some GM crops

Gaps in oversight of transgenic technologies allow scientists to test the waters for speciality varieties.

BY HEIDI LEDFORD

It took scientists 85 years to breed a commercial apple that could fend off apple scab, a devastating disease caused by the fungus *Venturia inaequalis*. In 1999, they finally produced a tasty variety that contained the *Vf* defence gene, bred in from an unappetizing relative. Instead of dousing orchards with fungicides 30 times a season, farmers could spray the resistant crop just twice.

But five years later, *V. inaequalis* had evolved and apples trees were becoming infected again. Breeders were back to square one. Even armed with modern breeding techniques and 15 known defence genes in the apple family, it would take another 40 years to breed a resistant strain conventionally, says Henk Schouten, a plant scientist at Wageningen University in the Netherlands.

So instead, Schouten has joined a small but growing pool of academics and companies hoping to taking advantage of the latest approaches in genetic engineering, while avoiding the lengthy and expensive burden of government regulation. Because he wants to insert DNA only from related apple varieties, Schouten argues that his product should not be regulated in the same way as genetically modified (GM) crops that are engineered with bacterial or viral DNA. Other pioneers argue that the techniques they are using to modify plants are safer than old technologies, and therefore do not need regulation. In some cases, US regulators have agreed.

Since 2010, the US Department of Agriculture has told at least 10 groups that their GM products would not require regulation (see 'Cropping out regulation') — removing a substantial financial barrier and speeding up development. That has encouraged academic labs and small companies to pursue speciality crops, such as apples, that have so far been ignored by biotechnology giants.

"There are any number of companies exploring new techniques to produce crops that don't trigger regulatory oversight," says Scott Thenell, managing director of Thenell & Associates, a consulting firm in Walnut Creek, California, that helps researchers to navigate GM-plant regulations. "And often, they are small or niche crops that can't support the escalating costs of regulatory approval."

The regulation of GM crops in the United States is based on laws that were not tailor-made for the technology. The Animal and Plant Health Inspection Service (APHIS), ▶

➲ NATURE.COM
For more on GM crops, see *Nature*'s special issue:
nature.com/gmcrops

the branch of the agriculture department responsible for overseeing GM crops, has so far stuck to a strict interpretation of a 1957 law designed to protect agriculture against plant pests that was co-opted in 1986 to regulate GM crops. At that time, GM crops were nearly always engineered using *Agrobacterium tumefaciens*, a bacterial pest that can insert DNA into plant genomes.

In 2011, APHIS regulators announced that a herbicide-tolerant Kentucky bluegrass would not fall under their purview, because the lawn-and-garden company developing it did not use *Agrobacterium* or any other plant-pest DNA to engineer the grass. The company, Scotts Miracle-Gro of Marysville, Ohio, instead used a gene gun to fire DNA-coated gold particles into plant cells. Some of that DNA is then incorporated into the genome.

For Greg Jaffe, director of biotechnology at the Center for Science in the Public Interest, a consumer advocacy group in Washington DC, the news highlighted the shortcomings of the US regulatory system for GM crops. "The whole system is a fiction," he says.

And some are starting to test the regulation-free waters. Scotts Miracle-Gro, for its part, has said that its bluegrass was not meant to be commercialized, and was just a test case to see how APHIS would respond. That is not the case for other groups that have been told that their GM products would not be regulated. Some include academic researchers, who are eager to avoid field-trial permits and special containment measures, and who want to encourage corporate development of niche crops.

Dennis Gray, a developmental biologist at the University of Florida in Apopka, is trying to use genes from grape varieties to engineer a wine grape that is resistant to Pierce's disease — a condition caused by a bacterium that has made it difficult to grow wine grapes in the state. He says that the lack of regulation is encouraging researchers like him to pursue such small-market crops. "Little agricultural labs just don't have access to the infrastructure and the money needed to move these forward."

Other emerging approaches may also escape regulation. Sally Mackenzie, a plant biologist at the University of Nebraska–Lincoln, contacted APHIS about the high-yield offspring of a transgenic sorghum grass plant — even though these offspring no longer contain the engineered gene. Mackenzie thinks that the transgene triggered an epigenetic change: it altered the plant's gene expression by changing the pattern of chemical groups added to its DNA rather than changing the DNA sequence itself. In 2012, APHIS regulators invited Mackenzie to the organization's headquarters in Riverdale, Maryland, and questioned her about this hypothesis. APHIS eventually notified her that it would not regulate her plants — a decision that Mackenzie says has accelerated her research and may allow her to launch a company to develop her grass variety.

Agricultural giants Monsanto, based in St Louis, Missouri, and Syngenta, headquartered in Basel, Switzerland, are vying to license the technology. "The first thing they asked me was, 'Have you been through APHIS?'" says Mackenzie.

Other companies are gauging their prospects with different DNA-modification tools, such as zinc-finger nucleases — enzymes that precisely target a region of the plant genome. In 2010, APHIS told Dow AgroSciences of Indianapolis, Indiana, that it would not regulate a herbicide-tolerant maize (corn) made using zinc-finger nucleases. Dow spokesman Garry Hamlin says that the company has since dropped the maize project, but is working with outside researchers to develop other crops using similar technology.

Jennifer Kuzma, a policy analyst at North Carolina State University in Raleigh, says that a lack of regulation for the latest approaches could fuel public suspicions about GM crops. "One could argue that the technologies are more targeted and you're doing things in a smarter way," she says. "The flip side is that they are so powerful you can engineer multiple genes at one time."

Not all companies are embracing the potential for freedom from regulation. Oliver Peoples, chief scientific officer at Metabolix, a plant-engineering company in Cambridge, Massachusetts, says that he would rather be regulated by APHIS to earn the public's trust. He notes that *Agrobacterium* inserts genes more efficiently than the gene-gun method. Although zinc-fingers are appealing for their specificity and their ability to escape regulation, companies do not yet have much experience in working with the technique, or navigating the patents needed to use it.

Schouten, meanwhile, did not skirt regulation for his apples after all. In April 2012, APHIS told him that the agency would regulate his variety in spite of the fact that the genes he introduced came from other apples. This was because he used *Agrobacterium* to insert the genes — it did not matter to regulators that no trace of *Agrobacterium* DNA remained in his plants.

Schouten is perplexed. If he had used a gene gun, he would have inserted DNA haphazardly and in a manner more likely to damage other sites in the genome — yet this remains the unregulated method. "To me, this is a very strange system," he says. ■

CROPPING OUT REGULATION

Since 2010, the US Department of Agriculture has told at least 10 groups that their genetically modified (GM) crops would not be regulated because a plant pest was not used to do the engineering.

Crop	Trait	Developer	Technique
Switchgrass	Easier conversion to biofuels	Ceres	Gene gun
Grapes	Red colour	University of Florida	Gene gun
Turf grasses	Herbicide tolerant	Scotts Miracle-Gro	Gene gun
Maize (corn)	Improved nutrition	Dow AgroSciences	Zinc-finger nuclease
Plums	Faster breeding	Appalachian Fruit Research Station	Non-transgenic offspring of GM parents
Tobacco	Faster breeding	North Carolina State University	Non-transgenic offspring of GM parents
Sorghum grass	Higher yields	University of Nebraska–Lincoln	Epigenetics
Not disclosed	Faster breeding	New Zealand Institute for Plant and Food Research	Non-transgenic offspring of GM parents
Ornamental plants	Not disclosed	BioGlow	Not disclosed
Not disclosed	Not disclosed	Cellectis	Meganuclease-targeted gene deletions

SOURCE: APHIS

MORE ONLINE

TOP STORY

XIAO YANG

Pesticide-resistance genes make weedy rice a worse pest go.nature.com/6swnpk

MORE NEWS

- Aquaculture farms cause subsidence in China's Yellow River delta go.nature.com/qmhtrc
- Measure of neural activity tracks waking states go.nature.com/dvfqll
- Climate change has made apples sweeter and softer go.nature.com/lzoikk

NATURE PODCAST

The 'sounds' of the stars; electric signals help plants heal; and turning skin cells into eggs and sperm nature.com/nature/podcast

EGG ENGINEERS

In a technical tour de force, Japanese researchers created eggs and sperm in the laboratory. Now, scientists have to determine how to use those cells safely — and ethically.

BY DAVID CYRANOSKI

Since last October, molecular biologist Katsuhiko Hayashi has received around a dozen e-mails from couples, most of them middle-aged, who are desperate for one thing: a baby. One menopausal woman from England offered to come to his laboratory at Kyoto University in Japan in the hope that he could help her to conceive a child. "That is my only wish," she wrote.

The requests started trickling in after Hayashi published the results of an experiment that he had assumed would be of interest mostly to developmental biologists[1]. Starting with the skin cells of mice *in vitro*, he created primordial germ cells (PGCs), which can develop into both sperm and eggs. To prove that these laboratory-grown versions were truly similar to naturally occurring PGCs, he used them to create eggs, then used those eggs to create live mice. He calls the live births a mere 'side effect' of the research, but that bench experiment became much more, because it raised the prospect of creating fertilizable eggs from the skin cells of infertile women. And it also suggested that men's skin cells could be used to create eggs, and that sperm could be generated from women's cells. (Indeed, after the research was published, the editor of a gay and lesbian magazine e-mailed Hayashi for more information.)

Despite the innovative nature of the research, the public attention surprised Hayashi and his senior professor, Mitinori Saitou. They have spent more than a decade piecing together the subtle details of mammalian gamete production and then recreating that process *in vitro* — all for the sake of science, not medicine. Their method now allows researchers to create unlimited PGCs, which were previously difficult to obtain, and this regular supply of treasured cells has helped to drive the study of mammalian reproduction. But as they push forward with the scientifically challenging transition from mice to monkeys and humans, they are setting the course for the future of infertility treatments — and perhaps even bolder experiments in reproduction. Scientists and the public are just starting to grapple with the associated ethical issues.

"It goes without saying that [they] really transformed the field in the mouse," says Amander Clark, a fertility expert at the University of California, Los Angeles. "Now, to avoid derailing the technology before it's had a chance to demonstrate its usefulness, we have to have conversations about the ethics of making gametes this way."

BACK TO THE BEGINNING

In the mouse, germ cells emerge just after the first week of embryonic development, as a group of around 40 PGCs[2]. This little cluster goes on to form the tens of thousands of eggs that female mice have at birth, and the millions of sperm cells that males produce every day, and it will pass on the mouse's entire genetic heritage. Saitou wanted to understand what signals direct these cells throughout their development.

Over the past decade, he has laboriously identified several genes — including *Stella*, *Blimp1* and *Prdm14* — that, when expressed in certain combinations and at certain times, play a crucial part in PGC development[3–5]. Using these genes as markers, he was able to select PGCs from among other cells and study what happens to them. In 2009, from experiments at the RIKEN Center for Developmental Biology in Kobe, Japan, he found that when culture conditions are right, adding a single ingredient — bone morphogenetic protein 4 (Bmp4) — with precise timing is enough to convert embryonic cells to PGCs[2]. To test this principle, he added high concentrations of Bmp4 to embryonic cells. Almost all of them turned into PGCs[2]. He and other scientists had expected the process to be more complicated.

➲ NATURE.COM
To see more on the creation of eggs from stem cells, visit:
go.nature.com/xuyljd

Saitou's approach — meticulously following the natural process — was in stark contrast to work that others were doing, says Jacob Hanna, a stem-cell expert at the Weizmann Institute of Science in Rehovot, Israel. Many scientists try to create specific cell types *in vitro* by bombarding stem cells with signalling molecules and then picking through the resulting mixture of mature cells for the ones they want. But it is never clear by what process these cells are formed or how similar they are to the natural versions. Saitou's efforts to find out precisely what is needed to make germ cells, to get rid of superfluous signals and to note the exact timing of various molecules at work, impressed his colleagues. "There's a really beautiful hidden message in this work — that differentiation of cells [*in vitro*] is really not easy," says Hanna. Harry Moore, a stem-cell biologist at the University of Sheffield, UK, regards the careful recapitulation of germ-cell development as "a triumph".

Until 2009, Saitou's starting point had been cells taken from a live mouse epiblast — a cup-like collection of cells lining one end of the embryo that forms at the end of the first week of development, just before the PGCs emerge. But to truly master the process, Saitou wanted to start with readily available, cultured cells.

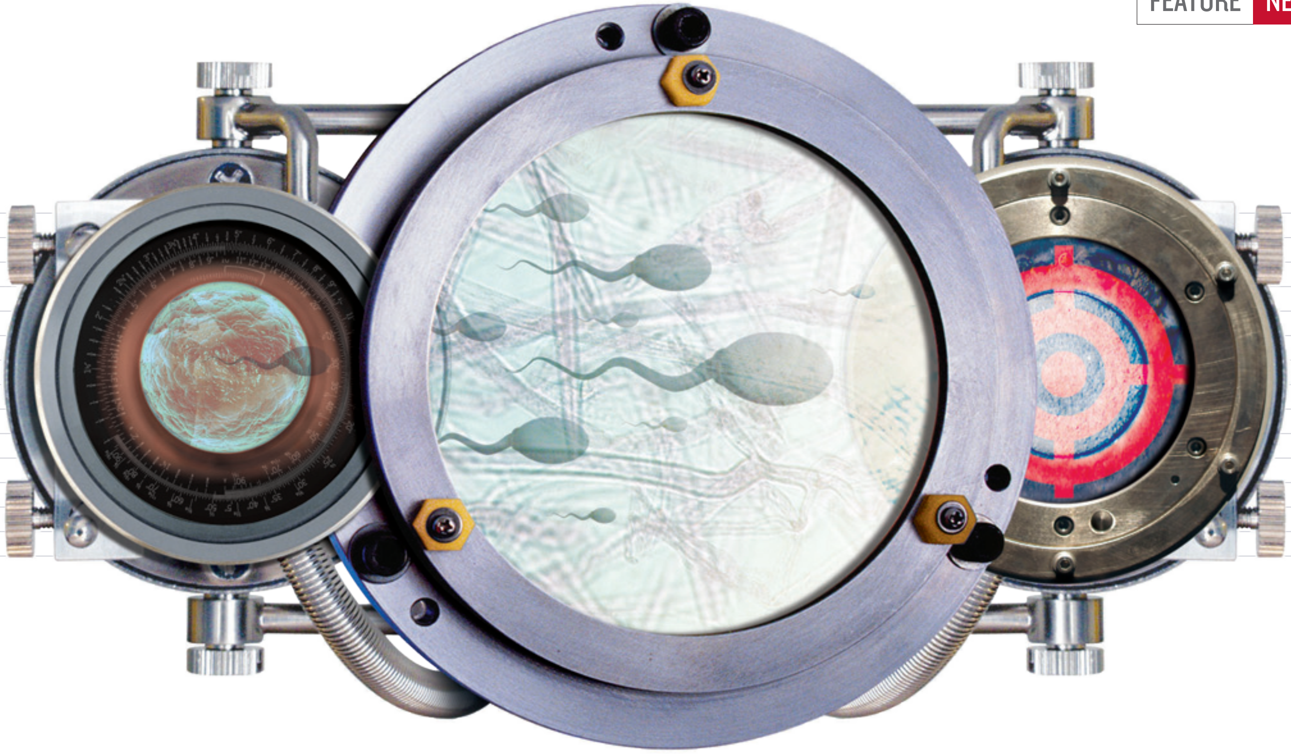


ILLUSTRATION BY VIKTOR KOEN

That was a project for Hayashi, who in 2009 had returned to Japan from the University of Cambridge, UK, where, like Saitou before him, he had completed a four-year stint in the laboratory of a pioneer in the field, Azim Surani. Surani speaks highly of the two scientists, saying that they "complement each other in temperament and in their style and approach to solving problems". Saitou is "systematic" and "single-minded about setting and accomplishing his objectives", whereas Hayashi "works more intuitively, and takes a broader view of the subject and has outwardly a more relaxed approach", he says. "Together they form a very strong team indeed."

Hayashi joined Saitou at Kyoto University, which he quickly found was different from Cambridge. There was much less time spent on theoretical discussions than Hayashi was used to; instead, one jumped into experiments. "In Japan we just do it. Sometimes that can be very inefficient, but sometimes it makes a huge success," he says.

Hayashi tried to use epiblast cells — Saitou's starting point — but instead of using extracted cells as Saitou did, he tried to culture them as a stable cell line that could produce PGCs. That did not work. Hayashi then drew on other research showing that one key regulatory molecule (activin A) and a growth factor (basic fibroblast growth factor) could convert cultured early embryonic stem cells into cells akin to epiblasts. That sparked the idea of using these two factors to induce embryonic stem cells to differentiate into epiblasts, and then to apply Saitou's previous formula to push these cells to become PGCs. The approach was successful[6].

To prove that these artificial PGCs were faithful copies, however, they had to be shown to develop into viable sperm and eggs. The

> **"THEY ARE SETTING THE COURSE FOR THE FUTURE OF INFERTILITY TREATMENTS."**

process by which this happens is complicated and ill understood, so the team left the job to nature — Hayashi inserted the PGCs into the testes of mice that were incapable of producing their own sperm, and waited to see whether the cells would develop[6]. Saitou thought that it would work, but fretted. "It seemed like a 50/50 chance," he says. "We were excited and worried at the same time." But, on the third or fourth mouse, they found testes with thick, dark seminiferous tubules, stuffed with sperm. "It happened so properly. I knew they would generate pups," says Hayashi. The team injected these sperm into eggs and inserted the embryos into female mice. The result was fertile males and females[6] (see 'Making babies').

They repeated the experiment with induced pluripotent stem (iPS) cells — mature cells that have been reprogrammed to an embryo-like state. Again, the sperm were used to produce pups, proving that they were functional — a rare accomplishment in the field of stem-cell differentiation, where scientists often argue over whether the cells that they create are truly what they seem to be. "This is one of the few examples in the entire field of pluripotent-stem-cell research where a fully functional cell type has been unequivocally generated starting from a pluripotent stem cell in a dish," says Clark.

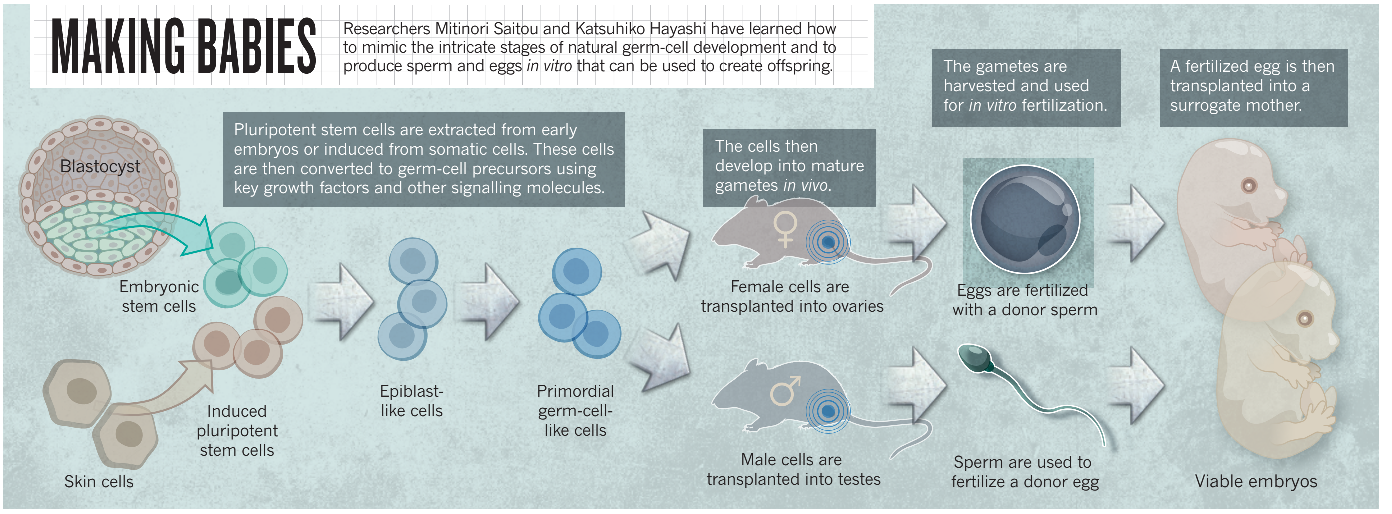
They expected eggs to be more complex, but last year, Hayashi made PGCs *in vitro* with cells from a mouse with normal colouring and then transferred them into the ovaries of an albino mouse[1]. The resulting eggs were fertilized *in vitro* and implanted into a surrogate. "I knew it had worked," he says, when he saw the pups' dark eyes pressing through their translucent eyelids.

GERM-CELL BOUNTY

Other researchers have been able to replicate the process to generate laboratory-grown PGCs (although none contacted by *Nature* had used them to produce live animals). Artificial PGCs are of particular use to scientists who study epigenetics: the biochemical modifications to DNA that determine which genes are expressed. These modifications — most often the addition of methyl groups to individual DNA bases — in some instances carry a sort of historical record of what an organism has experienced (for example, exposure to foreign chemicals in the womb). In a similar way to how they work in other cells, epigenetic markers push PGCs to their fate during embryonic development, but PGCs are unique because when they develop into sperm and eggs, the epigenetic markers are erased. This allows the cells to create a new zygote that is capable of forming all cell types.

Faults in subtle epigenetic changes are expected to contribute to infertility and the emergence of disorders such as testicular cancer. Already, Surani's and Hanna's groups have used the artificial PGCs to investigate the role of individual enzymes in epigenetic regulation, which may one day show how the epigenetic networks are involved in disease.

Indeed, the *in vitro*-generated PGCs offer millions of cells for scientists to study, instead of the 40 or so that can be obtained by dissecting



early embryos, says Hanna. "This is a big deal because here we have these rare cells — PGCs — that are undergoing dramatic genome-wide epigenetic changes that we barely understand," he says. "The *in vitro* model has provided unprecedented accessibility to scientists," agrees Clark.

CLINICAL RELEVANCE

But Hayashi and Saitou have little to offer to the infertile couples begging for their help. Before this protocol can be used in the clinic, there are large wrinkles to be ironed out.

Saitou and Hayashi have found that although the offspring generated by their technique usually seem to be healthy and fertile, the PGCs that these offspring generate in turn are not completely 'normal'. The second-generation PGCs often produce eggs that are fragile, misshapen and sometimes dislodged from the complex of cells that supports them[1]. When fertilized, the eggs often divide into cells with three sets of chromosomes rather than the normal two, and the rate at which the artificial PGCs successfully produce offspring is only one-third of the rate for normal *in vitro* fertilization (IVF). Yi Zhang, who studies epigenetics at Harvard Medical School in Boston, Massachusetts, and who has been using Saitou's method, has also found that *in vitro* PGCs do not erase their previous epigenetic programming as well as naturally occurring PGCs. "We have to be aware that these are PGC-like cells and not PGCs," he says.

In addition, two major technical challenges remain. The first is working out how to make the PGCs convert to mature sperm and eggs without transplanting them back into testes or ovaries; Hayashi is trying to decipher the signals that ovaries and testes give to the PGCs that tell them to become eggs or sperm, which he could then add to artificial PGCs in culture to lead them through these stages.

But the most formidable challenge will be repeating the mouse PGC work in humans. The group has already started tweaking human iPS cells using the same genes that Saitou pinpointed as being important in mouse germ-cell development, but both Saitou and Hayashi know that human signalling networks are different from those in mice. Moreover, whereas Saitou had 'countless' numbers of live mouse embryos to dissect, the team has no access to human embryos. Instead, the researchers receive 20 monkey embryos per week from a nearby primate facility, under a grant of ¥1.2 billion (US$12 million) over five years. If all goes well, Hayashi says, they could repeat the mouse work in monkeys within 5–10 years; with small tweaks, this method could then be used to produce human PGCs shortly after.

But making PGCs for infertility treatment will still be a huge jump, and many scientists — Saitou included — are urging caution. Both iPS and embryonic stem cells frequently pick up chromosomal abnormalities, genetic mutations and epigenetic irregularities during culture. "There could be potentially far-reaching, multi-generational consequences if something went wrong in a subtle way," says Moore.

Proof that the technique is safe in monkeys would help to allay concerns. But how many healthy monkeys would need to be born before the method could be regarded as safe? And how many generations should be observed?

Eventually, human embryos will need to be made and tested, a process that will be slowed by restrictions on creating embryos for research. New, non-invasive imaging techniques will enable doctors to sort good from bad embryos with a high degree of accuracy[7]. Embryos that seem to be similar to normal IVF embryos could get the go-ahead for implantation into humans. This might happen with private funding or in countries with less-restrictive attitudes towards embryo research.

When the technology is ready, even more provocative reproductive feats might be possible. For instance, cells from a man's skin could theoretically be used to create eggs that are fertilized with a partner's sperm, then nurtured in the womb of a surrogate. Some doubt, however, that such a feat would ever be possible — the Hinxton Group, an international consortium of scientists that discusses stem-cell ethics and challenges, concluded that it would be difficult to get eggs from male XY cells and sperm from female XX cells. "The instructions that the female niche is supplying to the male cell do not coordinate with each other," says Clark, a member of the consortium.

Saitou used iPS cells from male mice to create sperm and from female mice to create eggs, but he says that the reverse should be possible. If so, eggs and sperm from the same mouse could be generated and used for fertilization, producing something never seen before: a mouse created by self-fertilization. Neither Hayashi nor Saitou is ready to try this. "We would only do this [in mice] if there were a good scientific reason," says Saitou. Right now he does not see one.

The two scientists already feel some pressure from patients and Japanese funding agencies to move forward. The technique could be a last hope for women who have had no luck with IVF, or for people who had cancer in childhood and have lost the ability to produce sperm or eggs. Hayashi warns those who write to him that a viable infertility treatment could be 10 or even 50 years in the future. "My impression is that it is very far away. I don't want to give people unfeasible hope," he says.

Patients see the end result — success in mice — and often ignore the years of painstaking work that led to such a technical tour de force. They do not realize that switching from mice to humans means starting again almost from scratch, says Hayashi. The human early embryo is so different from the mouse that it is almost "like starting over on a process that took more than ten years". ■

David Cyranoski *is* Nature*'s Asia-Pacific correspondent.*

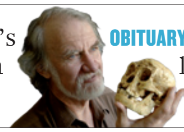
1. Hayashi, K. *et al. Science* **338,** 971–975 (2012).
2. Ohinata, Y. *et al. Cell* **137,** 571–584 (2009).
3. Saitou, M., Barton, S. C. & Surani, M. A. *Nature* **418,** 293–300 (2002).
4. Ohinata, Y. *et al. Nature* **436,** 207–213 (2005).
5. Yamaji, M. *et al. Nature Genet.* **40,** 1016–1022 (2008).
6. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. *Cell* **146,** 519–532 (2011).
7. Wong, C. C. *et al. Nature Biotechnol.* **28,** 1115–1121 (2010).

COMMENT

FICTION Margaret Atwood concludes her post-pandemic trilogy **p.398**

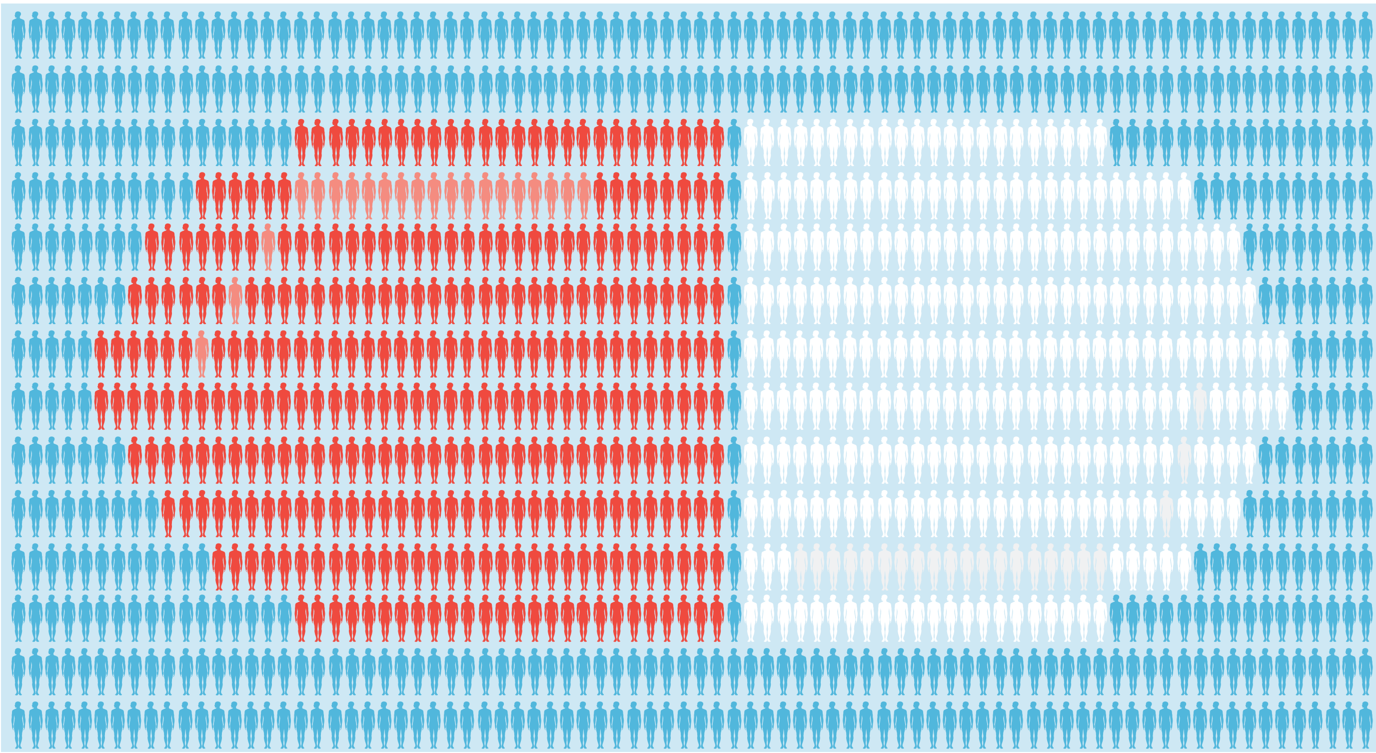
CONSERVATION Indigenous peoples in Panama need more say in forest plans **p.400**

HEALTH A congresswoman's call to limit antibiotics in farm animals **p.400**



OBITUARY Michael Morwood, leader of the 'Hobbit' fossil discovery **p.401**

CLAIRE WELSH/NATURE



Trial unpredictability yields predictable therapy gains

In decades of clinical-trial data, new treatments are better than standard ones just over half the time. That's as it should be, say **Benjamin Djulbegovic** and colleagues.

The effects of a new treatment are hard to predict. In the 1940s, pathologist Sidney Farber theorized that folate might help children with acute lymphoblastic leukaemia by stimulating blood cells; instead, he was surprised to find that the leukaemia cells proliferated. That 'failure' led him to try anti-folate drugs. These did lead to remissions, and were the first example of successful chemotherapy for cancer. In the 2000s, physicians thought that corticosteroids might help to reduce brain swelling following trauma, but randomized trials showed that they actually increased mortality. Surprisingly, however, corticosteroids were found to reduce the death rate in meningitis[1].

These are just a few examples of the steady therapeutic advances delivered over the past half century by the randomized controlled trial (RCT) system. But these improvements can feel frustratingly slow for patients and physicians. There is much soul searching in the drug-discovery community about how progress could be made more quickly[2].

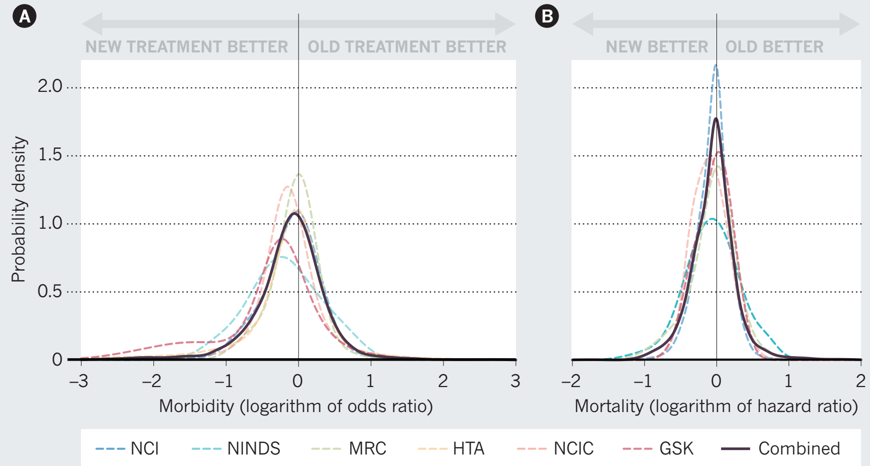
Here we provide empirical evidence that the system's success rate is optimal. We analysed hundreds of trials, published and unpublished, public and industry funded, involving hundreds of thousands of people over several decades. We find that just over half the time, RCTs show that new treatments are better than existing ones.

This success rate is incremental, but maintains a system that has served us well and is founded on the ethical and scientific necessity that the results of individual RCTs should not be predictable. We contend that the use of RCTs for assessing the effects of new treatments should not be ▶

fundamentally altered without analogous evidence that replacement systems will, on average, outperform them.

THE BEST MEDICINE

In just over 50% of randomized clinical trials, new treatments fare better than existing ones for both morbidity (A) and mortality (B).



NCI, US National Cancer Institute; NINDS, US National Institute of Neurological Disorders and Stroke; MRC, UK Medical Research Council; HTA, UK Health Technology Assessment Programme; NCIC, National Cancer Institute of Canada Clinical Trials Group; GSK, GlaxoSmithKline.

GENUINE UNCERTAINTY

Better drugs and therapies have come about because people participating in phase III trials are willing to be randomly allocated to new or existing treatments. Phase III trials are typically the final step in evaluating treatment efficacy. They are usually preceded by phase I trials that assess how a drug is metabolized, excreted and tolerated, and phase II trials that gather preliminary data on efficacy. Although phases I and II can occasionally identify new treatments with dramatic effects, thus obviating the need for further testing[3], phase III trials are usually required to judge whether new treatments are superior to existing ones.

On ethical as well as scientific grounds, RCTs should be done only when there are genuine uncertainties about the relative merits of alternative treatments[4]. If there were a high likelihood (say, more than 80%) that one of the treatments in a comparison was better than the other, it would be ethically unsound to deny some patients access to the superior treatment, and even if such a trial got past an ethics committee, well-informed patients would probably refuse to participate. In other words, if the results were predictable, the system of RCTs as we know it would cease[4]. Progress in therapeutics has occurred precisely because science and ethics require that the results of individual RCTs are not predictable.

Because this 'uncertainty requirement' — variously referred to as 'equipoise', 'the uncertainty principle' or 'the indifference principle'[4] — is insufficiently appreciated by the public, patients, research funders and investigators, we set out to test its long-term impact by calculating the average likelihood of a proposed new treatment being superior to established ones[5].

We conducted an analysis of 860 published and unpublished phase III RCTs performed by academics or pharmaceutical companies in six consecutive series of trials with a total of more than 350,000 patients: four series of 743 publicly sponsored trials over the past 50 years[6], and two series of 117 publicly and commercially sponsored clinical trials over the past 30 years[7] (see 'The best medicine'). Our results show that the probability of finding that a new treatment is better than a standard treatment is about 50–60%, confirming the theoretical predictions we made more than 15 years ago[4,5].

We found that in publicly sponsored RCTs, the likelihood that new treatments would work better than existing ones ranges from 57% to 63% for patient survival and from 55% to 66% for all primary outcomes (such as survival without recurrence of disease, response to treatment, symptom frequency and measures of disability). The only available comparable rates for industry-sponsored RCTs show that, overall, new treatments are superior to existing treatments for measures of morbidity (nausea, for example) in 75% of trials, but similar (53%) for survival[7]. Over time, the pattern in all trials has converged at around 50% (probably because earlier studies used inferior comparators) and applies across various clinical fields and types of treatment[6,7].

MAXIMUM GAIN

Philosophers of science have suggested that discovery in science happens most rapidly when only one or a few hypotheses are tested at a time[8]. The RCT system is paradigmatic of this approach. It has generated incremental advances that, together, translate into important improvements in health and lifespan. For example, five decades of controlled experimentation have seen cure rates for childhood leukaemia improve from 0% to more than 80% (ref. 6), yet in testing, only 2–5% of novel treatments have provided a breakthrough.

There is still room for improving existing practices for clinical trials. There is substantial avoidable waste in designing, conducting and reporting medical research[9]. For example, the results of only around 50% of RCTs are published — negative results and most industry trials remain hidden. The rigour of randomized trials can also be improved, for example by systematically taking into account all relevant previous research.

But our results show that the development of new treatments has been possible because the trials were done when unpredictability was greatest — in other words, when there was the most to gain[6,7]. The observed distribution of treatment successes is not an accident. There is a predictable relationship between the uncertainty requirement (the moral principle) on which trials are based and the outcomes of clinical trials[4].

In summary, our retrospective view of more than 50 years of randomized trials shows that they remain the 'indispensable ordeals' through which biomedical researchers' responsibility to patients and the public is manifested[10]. These trials may need a tweak and polish, but they're not broken. ■

Benjamin Djulbegovic *is professor in the Department of Internal Medicine, University of South Florida (USF), Tampa, and at the H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA.* **Ambuj Kumar** *is associate professor at the Department of Internal Medicine, USF and at the H. Lee Moffitt Cancer Center and Research Institute.* **Paul Glasziou** *is professor of evidence-based medicine at Bond University, Robina, Australia.* **Branko Miladinovic** *is assistant professor in the Department of Internal Medicine, USF.* **Iain Chalmers** *is coordinator of the James Lind Initiative, Oxford, UK.*
e-mail: bdjulbeg@health.usf.edu

1. Evans, I., Thornton, H., Chalmers, I. & Glasziou, P. *Testing Treatments: Better Research for Better Healthcare* 2nd edn (Pinter & Martin, 2011).
2. Micheel, C. M., Nass, S. J. & Omenn, G. S. (eds) *Evolution of Translational Omics: Lessons Learned and the Path Forward* (National Academies Press, 2012).
3. Glasziou, P., Chalmers, I., Rawlins, M. & McCulloch, P. *Br. Med. J.* **334,** 349–351 (2007).
4. Djulbegovic, B. *J. Med. Philos.* **32,** 79–98 (2007).
5. Chalmers, I. *Br. Med. J.* **314,** 74–75 (1997).
6. Djulbegovic, B. *et al. Cochrane Database Syst. Rev.* **10,** MR000024 (2012).
7. Djulbegovic, B. *et al. PLoS ONE* **8,** e58711 (2013).
8. Platt, J. R. *Science* **146,** 347–353 (1964).
9. Chalmers, I. & Glasziou, P. *The Lancet* **374,** 86–89, (2009).
10. Frederickson, D. S. *Control. Clin. Trials* **1,** 263–267 (1980).

SCIENCE FICTION

A post-pandemic wilderness

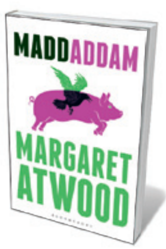
Paul McEuen relishes the final instalment of Margaret Atwood's sweeping trilogy about a dystopian world devastated by a 'hot bioform'.

A decade after Margaret Atwood began her great dystopian tale, we have at last reached the end of that road. The Canadian novelist has taken us from *Oryx and Crake* (2003) and *The Year of the Flood* (2009) to this final instalment, *MaddAddam*.

A global pandemic dominates the trilogy. In *Oryx and Crake*, a disillusioned bioengineer (Crake) unleashes a 'hot bioform' that kills most humans. *The Year of the Flood* revisits the pandemic through the lens of a religious cult called God's Gardeners, whose followers try to survive the ravages of the pathogen. *MaddAddam* completes the saga with the story of two members of the cult, Toby and Zeb, as they live through the aftermath of the plague. In the dystopian tradition, the trilogy is a window on our possible near future — in this case, one driven to disaster by human ingenuity gone wrong.

As *MaddAddam* opens, with almost all of humanity having perished in Crake's "Waterless Flood", it turns out that the bioengineer had good reason to reboot the human race. Atwood paints a picture of a pre-flood nightmare, class-divided, corporate and hegemonic. This was a world of *Hunger Games*-like death sports, rampant sexual enslavement and increasingly macabre genetically engineered hybrids. It begged to be wiped out.

The surviving humans must cope with a number of relics of pre-flood genetic tinkering. These include Pigoons — large, ferocious pigs with near-human intelligence, originally created for organ transplants — and domesticated goats with human hair known as Mo'Hairs. Also surviving is a small group of humanoids called Crakers, so-named for their creator and genetically modified to be polyamorous innocents with a predilection for eating kudzu (an invasive plant). These are the meek whom Crake would have had inherit the Earth, but they face many dangers. The remaining humans, especially Toby and Zeb, protect them from the Pigoons and a pair of murderous death-game survivors who have already raped and killed some of their clan.



MaddAddam
MARGARET ATWOOD
Bloomsbury Publishing: 2013.

As time passes, the Crakers begin to show signs of culture. They sing songs, beatify their now-dead creator, and hunger for more myths and stories about their origins. Toby, the book's main protagonist, provides these as best she can, and we watch with hope and dread as she spins child-like tales for the Crakers out of the unseemly facts of the Flood.

Many of these stories are told in flashback, particularly the full story of Adam and Zeb, who are in some ways the moral poles of *MaddAddam*. It is a biblical tale of grace and punishment, false idols and vengeance;

ILLUSTRATION BY JOHN RIORDAN

but Atwood keeps the morality multifaceted, making a case for both pacifism and, when absolutely necessary, murder.

Technology is the apple in the garden. In the pre-flood world, it evolved faster than it could be assimilated. Technology overwhelmed its creators, preying on their basest instincts and enslaving and degrading them. Plucked from the tree, it spread and destroyed.

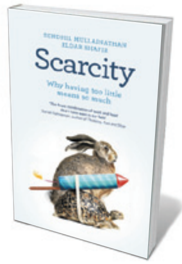
It is a pattern that threatens to repeat itself with the Crakers. Language, Atwood maintains, was humankind's first technology, and one of the most oddly chilling scenes occurs when the Crakers take the first bite of the apple. Toby is teaching one of the Crakers — a young boy named Blackbeard — about writing. The innocent Blackbeard refuses to accept the idea that pieces of the sensual world around him can be captured in lines on paper. Toby persists, showing the boy his name on a page. "This is how your name begins. B. Like bees. It's the same sound." But Blackbeard replies "That is not me," adding "It is not bees either."

Blackbeard learns in the end. He has tasted the fruit of the tree. But language is shown to be a saviour too. The secret to a new beginning for Toby, Zeb and the Crakers lies in forging deep links between the experiences of the humans and the Crakers, as well as the Mo'Hairs, bees and even Pigoons. This is how they start the world anew: as a process of weaving different languages and understandings of the world into a unified tapestry. Atwood shows us that what is missing in the fast-evolving technological world is a constant awareness of the link between the iPad and the exploited worker in China, or the hamburger on the plate and the factory-farmed cow.

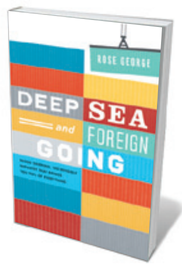
Will Atwood's imagined future be our own? Some elements of it will undoubtedly happen. Bioengineered meats are a staple in Atwood's pre-flood world, and earlier this month a bovine stem-cell hamburger created by Mark Post, a tissue engineer at Maastricht University in the Netherlands, was cooked and eaten. Will our technologies swallow us? The book's palindromic title suggests as much: disastrous ends yoked to new beginnings, with one flowing into the other in a never-ending cycle. But *MaddAddam* also tells us, even in the face of a disaster, to persevere. Atwood's book is a warning but also, in its final accounting, a hopeful meditation on the cycle of life, death and the possibility of life anew. ■

Paul L. McEuen *is the John A. Newman Professor of Physical Science at Cornell University, New York, and director of the Kavli Institute at Cornell for Nanoscale Science. His scientific thriller* Spiral *was named Best Debut Novel of 2012 by the International Thriller Writers association.*
e-mail: pmceuen@gmail.com

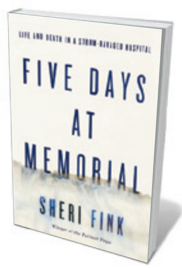
Books in brief



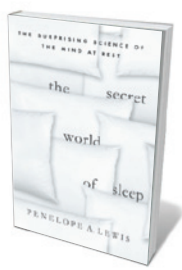
Scarcity: Why Having Too Little Means So Much
Sendhil Mullainathan and Eldar Shafir Allen Lane *(2013)*
Two scientists reveal that scarcity — "having less than you feel you need" — is a central factor in a raft of societal challenges. Economist Sendhil Mullainathan and psychologist Eldar Shafir posit that when we lack money or attention, for example, we obsess about it, leaving us little mental capacity to plan, meet other needs or practise self-control. We can become entrapped and eventually derailed by a vicious cycle. By reframing the dynamic as a mindset rather than a human failing, Mullainathan and Shafir train a new lens on chronic obesity, endemic poverty and desperate loneliness.



Deep Sea and Foreign Going: Inside Shipping, the Invisible Industry That Brings You 90% of Everything
Rose George Portobello Books *(2013)*
Some 746 million bananas ("one for every European") can fit into the largest container ship, notes journalist Rose George. About 100,000 cargo carriers ply the world's oceans, transporting 90% of our stuff. Yet these metallic Moby Dicks criss-crossing the lawless reaches of international waters can be hotbeds of crime, magnets for piracy and nemeses for sea life. Travelling with George on the Maersk Kendal from Felixstowe in the United Kingdom to Singapore, we are regaled — and horrified — by her salvos of facts. Riveting.



Five Days at Memorial: Life and Death in a Storm-Ravaged Hospital
Sheri Fink Crown *(2013)*
Medical ethics and disaster management take centre stage in this harrowing chronicle of a hospital besieged by Hurricane Katrina. Pulitzer-prizewinning journalist Sheri Fink tells how for five days in August 2005, a botched evacuation left hundreds trapped in the hot, increasingly filthy Memorial Medical Center in New Orleans. A handful of doctors and nurses were then alleged to have injected some of the severely ill with lethal drug doses. Fink reports on the ensuing nightmare with clarity and not a little compassion.



The Secret World of Sleep: The Surprising Science of the Mind at Rest
Penelope A. Lewis Palgrave Macmillan *(2013)*
The sleeping brain is not at rest: so reveals neuroscientist Penelope Lewis in this nippy primer on the biology and behaviour associated with snoozing. There is much to fascinate, such as the beneficial synaptic clear-outs enacted by slow-wave sleep, and the ascending reticular activating system — brainstem ganglia that send neurotransmitters to the rest of the brain to signal that it is time to wake up. From the latest on narcolepsy to the sleep-inhibiting qualities of smoked meat, this is wide-awake science.

Old Man River: The Mississippi River in North American History
Paul Schneider Henry Holt *(2013)*
It has been a bath for mammoths, a road for steamboats and a headache for engineers. The mighty Mississippi is a river that defines a nation, its tributaries branching out across the United States from Montana to Pennsylvania. In his natural and cultural history, Paul Schneider takes us from its origins 200 million years ago to its dammed and polluted present. His vast cast of heroes and eccentrics includes nineteenth-century showman Albert Koch, who haphazardly assembled fossils dug from Mississippi mud. **Barbara Kiser**

Correspondence

Antibiotics: support US policy change

As a microbiologist and member of the US Congress, I applaud your call to action on the overuse of antibiotics in agriculture (*Nature* **499,** 379, 394–396, 398–400; 2013). I have been fighting since 1999 to pass the Preservation of Antibiotics for Medical Treatment Act (PAMTA), which would ban the use of eight classes of medically important antibiotics in agriculture, with exceptions for treating sick animals.

In the United States, antibiotics are often distributed at sub-therapeutic doses to healthy farm animals to compensate for crowded and unsanitary living conditions or to promote growth.

In June, science ministers from the G8 nations discussed antibiotic resistance and committed to clamping down on overuse of antibiotics in health care, farming and fisheries. It is only through such coordinated international action that we can begin to hold back the tide of antibiotic-resistant bacteria.

Policy-makers need help, as you point out. I urge US readers to take a stand on this issue and ask their representatives and senators to co-sponsor PAMTA in the House of Representatives or the Preventing Antibiotic Resistance Act in the Senate (see www.louise.house.gov).

Louise M. Slaughter *Washington DC, USA.*
eric.walker@mail.house.gov

Antibiotics: collect more US data

The paucity of data on antibiotic use in livestock and poultry in the United States makes it hard for scientists to assess the relationship with antibiotic resistance (*Nature* **499,** 398–400; 2013). More comprehensive data need to be collected and made freely available to bring the United States in line with countries such as Denmark, where antibiotic use can be traced to individual animals.

The only US antibiotic data available are the sales figures that drug companies report to the Food and Drug Administration (FDA), which are published as total sales for each antibiotic class. Such broad aggregated data are of limited value, beyond confirming the extensive use of antibiotics in animals reared for food.

In February 2011, the Center for a Livable Future at the Johns Hopkins University in Baltimore, Maryland, and the Government Accountability Project (GAP) in Washington DC attempted to obtain more detailed antibiotics data from the FDA under the Freedom of Information Act. The FDA denied the request, claiming that these commercial data are confidential. In December 2012, GAP sued the FDA for access to the data (the case is ongoing).

Given that the misuse of antibiotics erodes their efficacy, there is an urgent need for greater transparency over their use. We contend that the FDA, as a public-health agency, is responsible to the public, not to the industry it regulates. It is imperative that more antibiotic data be released so that evidence-based public-health policies can be developed to combat antibiotic resistance.

Robert S. Lawrence, Keeve E. Nachman, Tyler J. Smith *Johns Hopkins Center for a Livable Future, Baltimore, Maryland, USA.*
tylsmith@jhsph.edu

Fukushima: unpaid soil-research effort

We and other particle physicists working voluntarily in Fukushima, Japan, where people were evacuated in 2011 because of the nuclear accident, are neither "opportunistic" nor "adventurous", as you quote (see *Nature* **499,** 265–266; 2013). We seek only to use our expertise to find a way to reduce the radioactive contamination of the area's soil.

We give up our weekends to work independently on analysing soil samples — a gruelling task in the winter months. We give lectures on radiation to those forced to leave their homes, so that they can better understand their plight and our efforts to remedy it. Our goal is to enable local farmers' livelihoods eventually to be restored.

Tokio Kenneth Ohska, Hiroshi Iwase *High Energy Accelerator Research Organization, Tsukuba, Japan.*
hiroshi.iwase@kek.jp

Fukushima: 'ecolab' branding insensitive

As the organizers of a symposium on the genetic effects of radiation following the Fukushima disaster — held at this year's annual meeting of the Society for Molecular Biology and Evolution — we object strongly to the headline of your report 'Fukushima offers real-time ecolab' (*Nature* **499,** 265–266; 2013).

In our view and those of many others from Japan and elsewhere who have communicated their objections to us, it conveys a lack of empathy among researchers for the suffering of the people and animals affected by the Fukushima power-plant disaster.

Scientists working on the consequences of the catastrophic events of March 2011, including the symposium panellists and ourselves, always take into primary consideration the pain of people in Fukushima. The researchers would never insult them by branding them or their natural environment as experimental material.

Your headline does not reflect the aims of our symposium or of the panellists' research

Tomoko Y. Steen *Georgetown University, Washington DC, USA.*
tys8@georgetown.edu
Marta L. Wayne *University of Florida, Gainesville, USA.*

Curb indigenous fears of REDD+

One of Panama's leading traditional indigenous authorities, the Guna General Congress, in June banned a project aimed at reducing emissions from deforestation and forest degradation (REDD+). The Congress, which controls about 7% of Panama's primary forests, went further, forbidding organizations in the Guna Yala territory from engaging in REDD+ activities, and walked out of REDD+ discussions. We believe that this crisis stems from a failure to build REDD+ capacity for indigenous people at all levels: it is time to pay more than lip service to their full and effective participation in REDD+.

REDD+ started well in Panama. The country put the rights of indigenous peoples on the agenda of the United Nations Framework Convention on Climate Change, and REDD+ project promoters complied with consent procedures of the Guna General Congress. Panama's National Coordinating Body of Indigenous Peoples (COONAPIP) drafted a plan in 2011 for comprehensive REDD+ capacity-building efforts in each indigenous territory. This would have stimulated debates about fears that REDD+ might threaten traditional land uses and rights, as well as possible ways forward. Knowledge transfer is the best antidote for the fear of REDD+.

The plan failed to receive UN funding. COONAPIP withdrew from the UN-REDD programme in February and called on indigenous peoples globally to proceed cautiously on REDD-related matters. If this fear of participation spreads beyond Guna Yala, the programme could be jeopardized in other Latin American countries.

Catherine Potvin, Javier Mateo-Vega *McGill University, Montreal, Canada; and Smithsonian Tropical Research Institute, Panama.*
catherine.potvin@mcgill.ca

Michael John Morwood

(1950–2013)

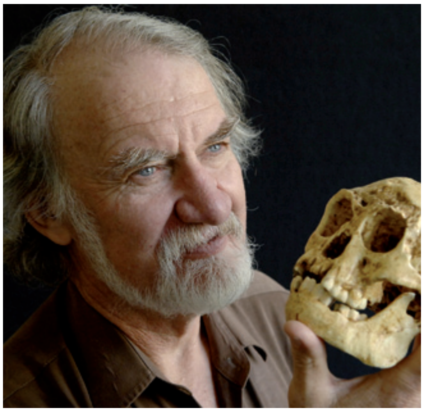
Rock-art archaeologist and driving force behind the 'Hobbit' discovery.

Michael John Morwood — Mike to his mates — was at heart what Australians would call a larrikin. Shaded by his battered bushman's hat on his frequent trips in the field, he wasted no time with small talk, and his quizzical stare gave him a slightly zealous demeanour. But his vision, intuition and leadership resulted in the 2003 discovery of *Homo floresiensis*, a species of archaic human identified from fossils found in eastern Indonesia. Given the type specimen's short stature, Morwood dubbed it 'Hobbit' after the fictional inhabitants of Middle-earth in J. R. R. Tolkien's *Lord of the Rings*.

Born in 1950 in Auckland, New Zealand, Morwood died of cancer in Darwin, Australia, on 23 July, en route to Indonesia. He was fascinated by Aboriginal rock art, the origins of the first Australians and their ancient connections with southeast Asia. As a state archaeologist in Queensland and during his PhD at the Australian National University in Canberra, Morwood pioneered studies that integrated rock art with artefacts recovered from excavations in Queensland. After joining the University of New England in Armidale, New South Wales, in 1981, he concentrated on sites in northern Queensland, culminating in the monograph *Quinkan Prehistory* (Tempus, 1995). This comprehensive work, edited with his long-time colleague Douglas Hobbs, gives a multidisciplinary perspective on 50,000 years of human activity in an environmental context, and set the tone for Morwood's subsequent projects in Western Australia and Indonesia.

Morwood served as president of the Australian Rock Art Research Association from 1992 to 2000, and in 2002 published *Visions from the Past* (Allen & Unwin). In this acclaimed book, he offers a continent-wide analysis of the rock art and archaeology of ancient Australia, drawing on his first-hand experiences spanning almost three decades. In 2007, he became a key figure in the nascent Centre for Archaeological Science at the University of Wollongong in New South Wales, where he planned and led further expeditions to Indonesia, and mentored the next generation of archaeologists.

In the mid-1990s, evidence of ancient contact between people of northern Australia and Indonesia led Morwood to launch a series of projects on Flores, an Indonesian island separated from mainland Asia by



OTAGO DAILY TIMES

several sea crossings. Morwood ventured first to central Flores, where in the 1960s the Dutch priest and amateur archaeologist Theodorus Verhoeven had contentiously reported finding 750,000-year-old stone tools. Morwood collaborated with researchers at Pusat Survei Geologi, the geological survey institute in Bandung, Indonesia, and with Australian geochronologists to prove that Verhoeven's conclusions were essentially correct. His team subsequently extended the antiquity of tool-making on Flores to 1 million years (possibly by ancestors of *H. floresiensis*) — the earliest evidence for humans east of Wallace's line, which separates the fauna of Asia and Australia.

Next, Morwood revisited another of Verhoeven's sites, a limestone cave in western Flores, looking for traces of ancestors of the first Australians. His direct and dogged style of negotiation with archaeologists at Pusat Arkeologi Nasional, the national archeological centre in Jakarta, succeeded where previous Australian archaeologists had failed, and he jointly led an Australian–Indonesian team to the cave in 2001. Two years later, after Morwood had returned to Java, leaving one of us (T.S.) to complete excavation of a 6-metre-deep hole at the site, Hobbit was discovered there unexpectedly. The news was relayed immediately to Morwood, who arranged safe transport of the fragile fossils to Jakarta for detailed study. A political and scientific saga unfolded, colourfully chronicled in Morwood and Penny van Oosterzee's popular-science book, *The Discovery of the Hobbit* (Random House, 2007). As other researchers jostled for the skeleton, bones were damaged, altered and sampled for DNA, resulting in acrimonious accusations and strenuous denials of incompetence and ethical misconduct.

Published in *Nature* in 2004, the discovery attracted intense scientific scrutiny and media coverage, propelling Morwood into the spotlight — sometimes reluctantly. That a 1-metre-tall human species with archaic features had survived until after *Homo sapiens* reached southeast Asia and Australia was a finding welcomed enthusiastically by some, but viewed sceptically by others. In response to concerns that *H. floresiensis* was not a new species, but a diseased member of *H. sapiens*, Morwood invited other human-evolution researchers to study and sample the fossils. This spirit of open enquiry epitomized his integrity and insistence on transparency. *H. floresiensis* is now generally accepted as a valid species, but its evolutionary lineage, geographical distribution and period of existence remain open questions that Morwood spent his final decade striving to answer.

Morwood prized his long-term collaborations in Indonesia and was always deeply respectful of indigenous communities and egalitarian in his dealings, treating senior colleagues and students alike. He took great care to educate, nurture and enthuse archaeology students and young researchers in Australia and Indonesia, and he inspired strong loyalties in his collaborators.

Field trips with Mike were memorable because of his insatiable curiosity and thirst for discovery (especially of another human species unknown to science), his whimsical sense of humour (Mike only half-joked that Hobbit should have been named *Homo hobbitus*) and his delight in adding a sword to his swashbuckling collection. He could also be exasperating: Mike paid scant attention to anything — or anyone — outside his immediate field of vision, and he had no patience for administrative paperwork that might impede his progress. But it was this single-minded drive and tenacity that enabled him to accomplish so much. We and many archaeologists in Australia and Indonesia will sorely miss Mike's camaraderie, restless energy and zest for adventure. ■

Richard G. Roberts *and* **Thomas Sutikna** *are at the Centre for Archaeological Science, University of Wollongong, Australia. T.S. is also at Pusat Arkeologi Nasional, Jakarta, Indonesia. They collaborated with Mike from the 1990s on projects in Australia and Indonesia, including the Hobbit discovery.*
e-mails: rgrob@uow.edu.au; thomasutikna@yahoo.com

NEWS & VIEWS

PLANT BIOLOGY

Electric defence

Herbivory and mechanical wounding in plants have been shown to elicit electrical signals — mediated by two glutamate-receptor-like proteins — that induce defence responses at local and distant sites. SEE LETTER P.422

ALEXANDER CHRISTMANN & ERWIN GRILL

The mammalian nervous system can relay electrical signals at speeds approaching 100 metres per second. Plants live at a slower pace. Although they lack a nervous system, some plants, such as the mimosa (*Mimosa pudica*) and the Venus flytrap (*Dionaea muscipula*), use electrical signals to trigger rapid leaf movements. Signal propagation in these plants occurs at a rate of 3 centimetres per second — comparable to that observed in the nervous system of mussels. On page 422 of this issue, Mousavi *et al.*[1] address the fascinating yet elusive issue of how plants generate and propagate electrical signals. The authors identify two glutamate-receptor-like proteins as crucial components in the induction of an electrical wave that is initiated by leaf wounding and that spreads to neighbouring organs, prompting them to mount defence responses to a potential herbivore attack.

As sessile organisms, plants have evolved diverse strategies to combat herbivores. These include mechanical defences, such as the thorns found on rose bushes, and chemical deterrents, such as the insect-neurotoxic pyrethrins of the genus *Chrysanthemum*. However, some plants do not invest in continuous defensive structures or metabolites, relying instead on the initiation of defence responses on demand[2]. This strategy requires an appropriate surveillance system and rapid communication between plant organs. A key player in orchestrating these reactions is the lipid-derived plant hormone jasmonate, which rapidly accumulates in organs remote from the site of herbivore feeding[3].

Mousavi *et al.* used thale cress (*Arabidopsis thaliana*) plants and Egyptian cotton leafworm (*Spodoptera littoralis*) larvae as a model of plant–herbivore interactions. The researchers placed the larvae on individual leaves and recorded changes in electrical potentials using electrodes grounded in the soil and on the surface of different leaves. The leaf-surface potential did not change when a larva walked on a leaf, but as soon as it started to feed, electrical signals were evoked near the site of attack and subsequently spread to neighbouring leaves at a maximum speed of 9 centimetres per minute. The relay of the electrical signal was most efficient for leaves directly above or below the wounded leaf. These leaves are well connected by the plant vasculature, which conducts water and organic compounds, and is a good candidate for the transmission of signals over long distances.

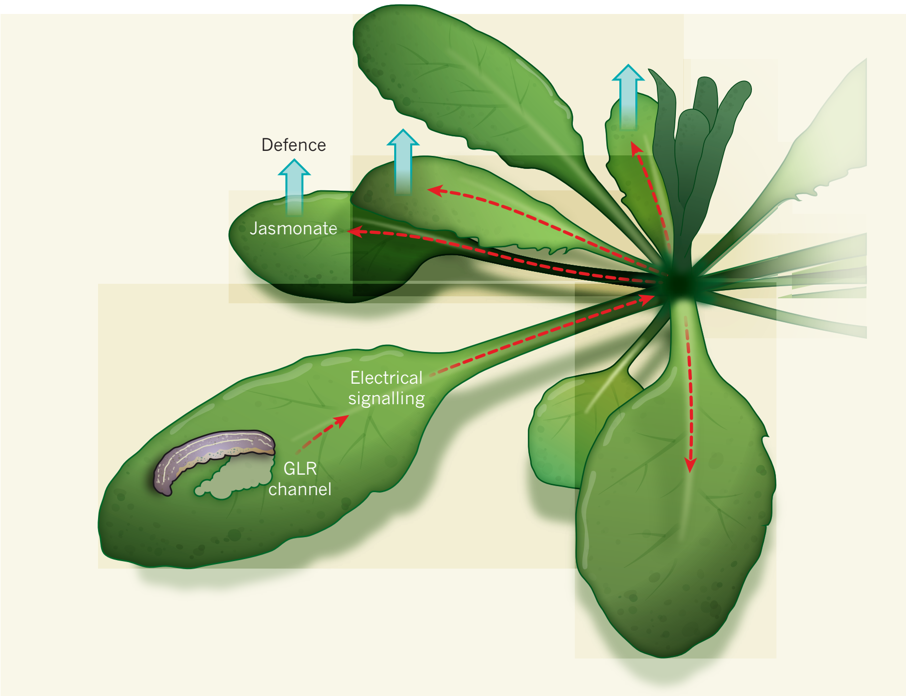


Figure 1 | Protective responses induced by electrical signalling. On herbivore attack, levels of the plant hormone jasmonate increase, triggering defence responses. Mousavi *et al.*[1] show that leaf injury, caused by herbivory or mechanical wounding, induces the transmission of electrical signals that are generated by the activity of glutamate-receptor-like (GLR) ion channels. These signals induce the formation of jasmonate at local and distant sites in the plant.

At all sites that received the electrical signals, jasmonate-mediated gene expression was turned on and initiated defence-responsive gene expression. In a mutant *A. thaliana* plant lacking the receptor for jasmonate, an electrical signal was propagated but no defence response was elicited. Defence responses also failed to occur at remote sites when the transmission of the electrical signal was prevented by ablation of the damaged leaf before the signal had passed the leaf stalk. These fascinating observations clearly demonstrate that electrical signal generation and propagation have a crucial role in the initiation of defence responses at remote sites upon herbivore attack.

The salivary secretions of herbivores contain elicitor molecules that are recognized by the host plant[4,5] and that induce jasmonate-mediated defence responses. However, Mousavi and colleagues found that extensive mechanical wounding (in the absence of herbivory) also initiated electrical signal transmission and jasmonate biosynthesis. In addition, a herbivore-response gene-expression pattern could be artificially induced by applying electric pulses that mimicked the plant's electrical signals. Thus, it remains unclear how the electrical signals are interpreted to stimulate jasmonate biosynthesis.

The authors next investigated which cellular components are involved in generating the electrical signals, by screening *A. thaliana* plants defective in candidate ion pumps and

channels. They found that loss of function of certain members of the glutamate-receptor-like (GLR) family of ion-channel proteins — some of which form calcium-ion-permeable channels that can be activated by agonists such as glutamate and serine[6,7] — affected wound-induced signal generation. Indeed, combined disruption of the genes encoding two of these channels, *glr3.3* and *glr3.6*, resulted in the electrical wave no longer propagating after wounding.

"Electrical signals evoked near the site of attack spread to neighbouring leaves at a maximum speed of 9 centimetres per minute."

Thus, it seems that herbivory and mechanical wounding trigger the local generation of an electrical signal through the activity of GLRs; this signal then spreads to neighbouring organs where the biosynthesis of jasmonate is induced, in turn triggering jasmonate-dependent defence responses (Fig. 1). Several questions emerging from this study will foster future research efforts. For example, how do feeding and mechanical wounding activate the GLRs? Might calcium ions be involved in the generation and maintenance of the electrical wave? It will also be intriguing to elucidate whether GLRs relay the faster electrical signalling that triggers movement in mimosa and the Venus flytrap.

Plant wounding is also known to evoke an extracellular wave of reactive oxygen species (ROS), which propagates at a speed[8] comparable to that recorded by Mousavi *et al.* for the electric signals. But the authors found that inhibiting wound-induced ROS generation did not substantially disrupt electric signalling, so it remains to be determined whether there is an interaction between wound-induced ROS signalling and electric signalling.

It is interesting to note that plant GLRs are structurally related to vertebrate ionotropic glutamate receptors, which are important for rapid excitatory synaptic transmission in the nervous system. Insect feeding on leaves has also been shown to generate an electric wave by a continuous relay of cell-membrane depolarizations[4] that is reminiscent of excitatory signal propagation in animals. Together, these findings imply that ionotropic glutamate-receptor-type proteins must have existed before animals and plants diverged. These ancestral proteins might already have functioned in the generation of long-distance warning signals to elicit the timely initiation of protective responses. ■

Alexander Christmann *and* **Erwin Grill** *are in the Department of Plant Sciences, Life Science Centre Weihenstephan, Technische Universität München, D-85354 Weihenstephan, Germany.*
e-mails: alexander.christmann@wzw.tum.de; erwin.grill@wzw.tum.de

1. Mousavi, S. A. R., Chauvin, A., Pascaud, F., Kellenberger, S. & Farmer, E. E. *Nature* **500,** 422–426 (2013).
2. Meldau, S., Erb, M. & Baldwin, I. T. *Ann. Bot.* **110,** 1503–1514 (2012).
3. Howe, G. A. & Jander, G. *Annu. Rev. Plant Biol.* **59,** 41–66 (2008).
4. Maffei, M., Bossi, S., Spiteller, D., Mithöfer, A. & Boland, W. *Plant Physiol.* **134,** 1752–1762 (2004).
5. Dinh, S. T., Baldwin, I. T. & Galis, I. *Plant Physiol.* **162,** 2106–2124 (2013).
6. Vincill, E. D., Bieck, A. M. & Spalding, E. P. *Plant Physiol.* **159,** 40–46 (2012).
7. Michard, E. *et al. Science* **332,** 434–437 (2011).
8. Miller, G. *et al. Sci. Signal.* **2,** ra45 (2009).

ASTROPHYSICS

Twinkling stars

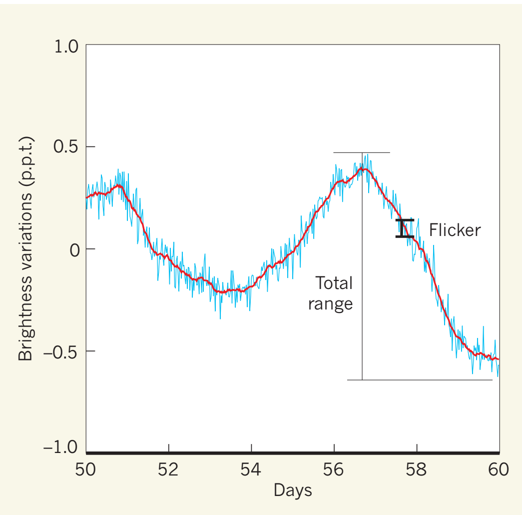
A correlation between stellar brightness variations and the gravitational acceleration at a star's surface has been observed that allows this acceleration to be measured with a precision of better than 25%. **SEE LETTER P.427**

JØRGEN CHRISTENSEN-DALSGAARD

"Twinkle, twinkle little star, how I wonder what you are." Given the wording of this old nursery rhyme, it is highly satisfying that Bastien *et al.*[1] (page 427 of this issue) find that a star's twinkle may hold the key to determining its properties. The authors used data from NASA's Kepler space mission to show that accurate measurements of variations in a star's light reveal information about the acceleration of gravity at the star's surface. This result is significant for the characterization of stars, and in particular for the determination of radii of stars hosting planetary systems.

Essentially all knowledge about distant stars derives from observation of the light emitted by their outer layers. Therefore, the properties of these layers are central to the study of stars. These properties have conventionally been obtained from analysis of stellar spectra, but the gravitational acceleration (g) has proved notoriously difficult to nail down, and the resulting uncertainty about this quantity has substantial effects on the measurement of other properties, such as temperature and chemical composition.

Analyses of variations in stellar brightness caused by stellar oscillations (asteroseismology), particularly those based on the spectacular data from the Kepler mission[2], provide precise determinations of g but require extensive observations and complex analysis, which are available for only a limited number of stars. However, stellar oscillations are not the only factor that contributes to variations in brightness. Bastien and colleagues show that g is also reflected in these variations.



REF. 1

Figure 1 | Stellar variability. Bastien *et al.*[1] describe the brightness variations of Sun-like stars in terms of variations on timescales of days (total range; here about 1 part per thousand) and of variations on timescales shorter than 8 hours (flicker; here roughly 0.03 p.p.t.). The red curve shows the result of smoothing the blue curve with an 8-hour running mean.

One of the properties of a star's brightness variations measured by Bastien *et al.* is the total range of the variations. This includes variations on timescales of days that may have a number of causes, such as the rotation of large starspots across the disc of the star. In addition to this total range, the authors characterize the variations in terms of what they call 'flicker' — variations that occur on timescales shorter than eight hours (Fig. 1). In the Kepler data, they identify a substantial fraction of Sun-like stars that have a low range, defining what they dub 'flicker floor' (see Fig. 3 in the paper). Bastien *et al.* find that, for a subset of these flicker-floor stars whose precise values of g are known from asteroseismology[2], there is a close correlation between flicker and g, with the amplitude of the flicker increasing with decreasing g. For other stars on

the flicker floor, this correlation provides a means of determining *g* with a precision of better than 25% — between two and three times better than the precision obtained from conventional observations.

As Bastien *et al.* note, the general dependence of brightness variations on stellar properties can be described in terms of stellar evolution. Young stars tend to have stronger magnetic activity, with many starspots, and hence display a large total range of variations. With increasing age, this activity diminishes and the stars settle on the flicker floor. As the stars grow older their radii increase, leading to a lower *g* and a higher flicker.

Bastien and colleagues demonstrate how flicker can be used to measure *g*, but do not provide a detailed analysis of the physical nature of the flicker. The stars for which the investigation was carried out have outer convection zones, in which energy is transported to the surface through the motion of gas. In the Sun, this transport is visible in granulation — a time-varying pattern of small-scale brighter and dimmer regions on the solar surface that reflects hot, rising and cooler, sinking gas pockets. Granulation also leads to minute variations in the total solar brightness.

The authors' study indicates that stellar granulation is a contributor to flicker. Indeed, the spatial scale and other properties of granulation depend on *g* (ref. 3), with lower *g* resulting in a larger scale and thus probably causing larger brightness variations on timescales relevant to flicker, in agreement with the correlation that the authors found. Further support for the relationship between granulation and flicker comes from other Kepler observations and modelling of red-giant stars[4]. Brightness variations caused by granulation are expected in all the stars considered by the authors, hence defining a lower limit to the variations — the flicker floor. A better physical understanding of the origin of flicker might allow the observed brightness variations to be used to probe the dynamics of the outermost stellar layers. The resulting improved stellar modelling could, in turn, improve the accuracy with which *g* can be determined.

> ***"Studying the twinkling of stars does indeed help us to understand what they are."***

The very small amplitude of flicker makes it essentially unobservable using ground-based telescopes, owing to the effect of Earth's atmosphere. However, the ability to measure flicker with Kepler observations will be valuable in the continuing analysis of Kepler data on exoplanets, which are detected through the slight dimming of a star's light as a planet transits, or passes in front of it. An accurate determination of *g* from flicker greatly aids the analysis of spectroscopic observations used to infer the chemical composition of planet-hosting stars, and so advances our understanding of planet formation[5]. Furthermore, planetary transit observations provide a measurement of only planetary radius relative to stellar radius, and uncertain information about stellar radii hampers the characterization of the planets. With knowledge of *g* from flicker, as well as of the surface temperature and composition of the star, fits of stellar models to these quantities can be used to obtain a more precise value of the stellar radius, and hence of the planetary radius.

Beyond Kepler, the authors' technique will be valuable for NASA's planned Transiting Exoplanet Survey Satellite (TESS), which is slated for launch in 2017. TESS will carry out an all-sky survey for extrasolar planetary systems by monitoring at least half a million stars, and will require efficient methods to characterize the target stars. The same applies to the European Space Agency's Planetary Transits and Oscillations of Stars (PLATO) exoplanet mission, should it be selected for launch in 2022–24.

Therefore, Bastien and colleagues' analysis holds great promise for measuring stellar properties and understanding the complex dynamics of the outermost layers of stars. Studying the twinkling of stars does indeed help us to understand what they are. ■

Jørgen Christensen-Dalsgaard *is in the Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, 8000 Aarhus C, Denmark.*
e-mail: jcd@phys.au.dk

1. Bastien, F. A., Stassun, K. G., Basri, G. & Pepper, J. *Nature* **500,** 427–430 (2013).
2. Chaplin, W. J. *et al. Science* **332,** 213–216 (2011).
3. Trampedach, R., Asplund, M., Collet, R., Nordlund, Å & Stein, R. F. *Astrophys. J.* **769,** 18 (2013).
4. Mathur, S. *et al. Astrophys. J.* **741,** 119 (2011).
5. Buchhave, L. A. *et al. Nature* **486,** 375–377 (2012).

BIOTECHNOLOGY

Programming genomes with light

The combination of two techniques — optogenetics and genome editing using engineered nucleases — now provides a general means for the light-controlled regulation of any gene of interest. SEE LETTER P.472

ANDREAS MÖGLICH & PETER HEGEMANN

In this issue, Konermann *et al.*[1] combine two sparkling biological technologies developed over the past decade. The first is optogenetics[2], the process by which light-responsive proteins are engineered into target cells and used to regulate their activity. The second is the use of sequence-targeted DNA-cleaving enzymes to specifically alter the genome. By uniting these techniques, the authors present a versatile method for targeted control of gene transcription and genomic modifications.

Optogenetics can be used in cells and in living organisms, and allows cellular regulation using light of different colours, intensities and duration in a graded, non-invasive, reversible and spatiotemporally precise fashion. Most optogenetic applications so far have relied on the use of light-sensitive ion channels and ion pumps to modulate the voltage dynamically across biological membranes, in particular to elicit action potentials in neurons.

Nature offers a plethora of other processes that are regulated by light, such as those controlled by photoreceptors in plants and microorganisms. With only a few exceptions, these processes are intricately tied to their organism of origin, and their deployment in others is challenging. However, their existence suggests that optogenetics could be extended to regulating enzyme activity or could be used to induce more persistent effects by targeting DNA. Indeed, natural photoreceptors have provided design blueprints for the engineering of several biological systems with customized light responses[3].

A particularly versatile strategy uses photoreceptors that associate with other proteins in a light-regulated process. In terms of performance, robustness, response time and ease of use, the blue-light-responsive protein cryptochrome 2 and its light-induced interaction with its partner protein CIB1 (ref. 4) currently have the edge over alternative photodimerizers such as the red-light-responsive phytochrome–PIF pair[5]. Several laboratories have successfully modulated gene transcription using photodimerizing proteins[4–8]. Initially, these optogenetic systems were directed to specific DNA sites by coupling to the DNA-binding part of the transcriptional-activator

protein Gal4 (ref. 5). Light exposure recruited an interacting protein that was coupled to the activation domain of Gal4, thus initiating transcription. Although these systems are powerful[7,8], they are inherently limited because they use DNA-binding domains with fixed target-sequence specificity, and because target genes have to be introduced into the host genome as exogenous DNA templates.

In parallel with the introduction of optogenetics, DNA-engineering strategies have been developed that can target unique sites among the billions of nucleotides in a genome. Early versions of such approaches[9] were based on zinc-finger and transcription-activator-like effector (TALE) proteins, which contain repetitive amino-acid sequences that recognize single DNA nucleotides or nucleotide triplets, and introduce double-stranded DNA breaks on binding to these sequences. The DNA-repair process that is activated in response to this damage can be used to introduce novel genetic elements at the site. However, adjusting the sequence specificity of zinc-finger and TALE proteins entails the laborious production of customized proteins.

A more recently developed approach, called the CRISPR–Cas system[10,11], overcomes this limitation. In this system, an endonuclease enzyme that induces a double-strand break is used with a sequence-specific guide RNA molecule — simply replacing the guide RNA is sufficient for sequence adaptation. The CRISPR–Cas technology stands to make engineering of zinc-finger and TALE proteins obsolete and to render genome engineering fast, efficient and inexpensive.

Capitalizing on their expertise in both optogenetics and genome engineering, Konermann *et al.* have overcome the sequence-restriction problem of earlier light-activated transcription-modulation approaches in their light-inducible transcriptional effector (LITE) system (Fig. 1a). The system uses a TALE protein coupled to cryptochrome 2, and CIB1 coupled to the transcriptional-activator protein VP64. This combination results in the cellular transcriptional machinery being recruited to the genomic site defined by the TALE protein when blue light is absorbed. The authors showed *in vitro* that, following light exposure, site-specific gene expression was enhanced by 10–20 times compared with darkness, and they convincingly validated the technology in mouse neurons and in the brains of conscious mice by monitoring light-mediated transcription of the genes *Grm2* and *Neurog2*.

The LITE approach has several favourable characteristics. First, the light-responsive molecules of cryptochrome 2 are the chromophores flavin-adenine dinucleotide and methyltetrahydrofolate, which are universally abundant. Second, induction of transcription occurs within minutes of light exposure.

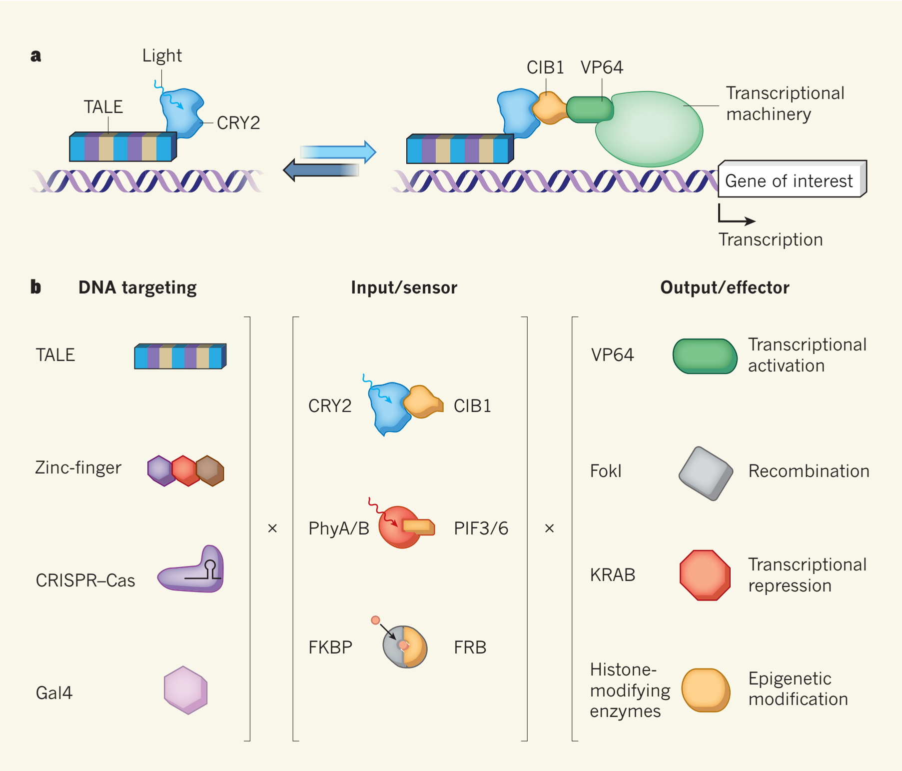


Figure 1 | Modular control of genome function. **a**, Konermann and colleagues' LITE system[1] uses transcription-activator-like effector (TALE) proteins — which specifically bind to unique DNA sequences — coupled to the light-sensitive photoreceptor cryptochrome 2 (CRY2). On light exposure, a complex of CRY2's interaction partner CIB1 coupled to the transcriptional-activator protein VP64 is attracted to CRY2, and VP64, in turn, attracts the cellular transcriptional machinery, initiating transcription at the target site. This reaction is reversible. **b**, The system is highly versatile because the various components can be interchanged. DNA targeting can be achieved using Gal4, zinc-finger proteins or the CRISPR–Cas system. Proteins that respond to light of different colours (such as red light for the phytochromes A/B (PhyA/B) and their interaction partners PIF3/6) or small molecules (such as rapamycin for the FKBP–FRB interaction pair) can be used as the sensor and recruitment molecules. Also, different output molecules can be used for various effects, including recombination (using endonucleases such as FokI), transcriptional repression (through the protein KRAB[11]) or histone modification (using enzymes that elicit epigenetic effects).

Third, the response can be graded with light dose and is fully reversible after light retraction. Finally, because light can be applied non-invasively, its use is not restricted to cultured cells but extends to freely moving animals, as established for conventional optogenetic tools[2].

Great power lies in the modularity and resultant versatility of this technique (Fig. 1b). By replacing constituent modules of LITE, the system can be tuned to be sensitive to light of different colours or to have different effector outputs. The authors impressively demonstrated this second possibility by interfacing LITE with various molecules that modify histones — the proteins around which DNA is wrapped. They show that their system can be used to site-specifically enhance histone methylation and acetylation — two epigenetic modifications that regulate the rate of gene transcription. The LITE approach thus enriches the optogenetic arsenal with novel applications.

Similarly, the TALE module of LITE can be exchanged for other DNA-binding modules, including ones based on the CRISPR–Cas system, as Konermann *et al.* demonstrate. Because the CRISPR–Cas system can be rapidly directed to different DNA sites, this will allow faster fine-tuning of the efficacy of any LITE experiments. Thus, the combination of CRISPR–Cas and LITE may truly usher in a new era of systems biology, in which gene expression and epigenetic modifications can be manipulated at the genome level with supreme sequence specificity, exquisite temporal resolution and full reversibility.

As with any new technology, there is room for improvement. In particular, it would be desirable to increase the degree of transcriptional activation by LITE. There is also the question of where the LITE system should be positioned in the genome to achieve maximum effect, but this can be easily addressed with the rapid manipulation offered by the CRISPR–Cas system. Even in its present implementation, LITE represents a powerful approach to light-controlled genome programming. Given its versatility, ease of use, performance and potential for automation, we expect this

technology to be widely and rapidly taken up across many biological disciplines. ■

Andreas Möglich *is in the Department of Biophysical Chemistry at the Institute of Biology, Humboldt-Universität zu Berlin, 10115 Berlin, Germany.* **Peter Hegemann** *is in the Department of Experimental Biophysics at the Humboldt-Universität zu Berlin.*
e-mails: hegemape@rz.hu-berlin.de; andreas.moeglich@hu-berlin.de

1. Konermann, S. *et al. Nature* **500,** 472–476 (2013).
2. Yizhar, O., Fenno, L. E., Davidson, T. J., Mogri, M. & Deisseroth, K. *Neuron* **71,** 9–34 (2011).
3. Moffat, K., Zhang, F., Hahn, K. M. & Möglich, A. in *Optogenetics* (eds Hegemann, P. & Sigrist, S.) 7–22 (Walter de Gruyter, 2013).
4. Kennedy, M. J. *et al. Nature Meth.* **7,** 973–975 (2010).
5. Shimizu-Sato, S., Huq, E., Tepperman, J. M. & Quail, P. H. *Nature Biotechnol.* **20,** 1041–1044 (2002).
6. Wang, X., Chen, X. & Yang, Y. *Nature Meth.* **9,** 266–269 (2012).
7. Polstein, L. R. & Gersbach, C. A. *J. Am. Chem. Soc.* **134,** 16480–16483 (2012).
8. Müller, K. *et al. Nucleic Acids Res.* **41,** e124 (2013).
9. Boch, J. *et al. Science* **326,** 1509–1512 (2009).
10. Charpentier, E. & Doudna, J. A. *Nature* **495,** 50–51 (2013).
11. Gilbert, L. A. *et al. Cell* **154,** 442–451 (2013).

CONDENSED-MATTER PHYSICS

A solid triple point

The observation of a triple point of coexistence between two insulating phases and a conducting phase in vanadium dioxide reveals physics that may help to unravel the role of electronic correlations in this material. SEE LETTER P.431

DOUGLAS NATELSON

Much of condensed-matter physics is concerned with thermodynamic phases, their properties and their transitions. In correlated materials, the electron–electron and electron–lattice interactions result in a competition between various electronic, magnetic and structural phases. The transitions between competing phases can reveal information about the underlying states that is otherwise difficult to obtain. On page 431 of this issue, Park *et al.*[1] use a micromechanical device and single-crystal nanobeams to determine with high precision the tensile stress–temperature phase diagram of vanadium dioxide (VO_2), an archetypal correlated oxide. Their experiment reveals a surprising and interesting fact: the metal–insulator phase transition for which VO_2 is famous is in fact a triple point, a rare circumstance in which three phases (here two insulators and a metal) can coexist. The experiment also determines the entropy differences between the various phases — information crucial to a complete understanding of the transition.

When a large amount of a substance (such as water) is brought together, it may exist in distinct phases (such as solid, liquid and gas). At given conditions, for example at a particular pressure and temperature, the thermodynamically stable phase is the one with the lowest free energy, which is determined by the arrangement, motion and interactions of the constituents. A phase diagram is a map of the stable phases as a function of parameters such as pressure and temperature.

When two phases coexist stably, their free energies must be equal, and for a single species, this condition leads to a coexistence 'line' for the two phases in the phase diagram. For example, ice and liquid water coexist in equilibrium at 0 °C and atmospheric pressure, and increasing the pressure decreases the melting point. Similarly, liquid water and water vapour coexist stably at 100 °C and atmospheric pressure, with decreasing pressure reducing that temperature (hence water boils at a lower temperature on top of a mountain than at sea level). These two coexistence curves can intersect only at a single value of pressure and temperature — a triple point (Fig. 1a). For water, this happens at 0.01 °C and 612 pascals. This particular triple coexistence defines the Kelvin temperature scale[2].

In VO_2, the competing phases of interest are all solids, albeit with different lattice structures and electronic properties: a high-temperature metallic phase (with a rutile lattice structure, R), and two insulating phases (with monoclinic structures, M1 and M2). The competition between these phases is of great interest because of the marked change in electronic and optical properties that occurs at the metal–insulator transition, the proximity of the transitions to room temperature, and the need to better understand the underlying physics. The relative importance of electron–electron interactions (Mott physics) and lattice distortion (Peierls physics) in stabilizing the M1 phase has been debated for decades. In addition to temperature, the intensive quantity relevant to VO_2 is the tensile stress rather than the pressure. Controlling this stress makes measurements in bulk crystals and thin films challenging.

Single-crystal VO_2 nanobeams[3] with a well-defined tensile-stress profile along the beam have been a boon to those trying to understand the intrinsic physics of this material. Park *et al.* attached an individual single-crystal VO_2 nanobeam to bridge a notched silicon structure, and used a piezo actuator to apply a controlled longitudinal deformation to the nanobeam, and so vary its length. Through polarized optical microscopy, Raman microscopy and electrical measurements, they identified regions of the suspended beam in the M1, M2 and R phases. Because the entire system was mounted on a temperature-controlled stage, the authors were able to determine the tensile stress–temperature phase diagram (Fig. 1b) by performing measurements of phase composition as a function of temperature at fixed length (which they can relate to the stress) and as a function of length at fixed temperature. To obtain measurements at zero stress, they broke

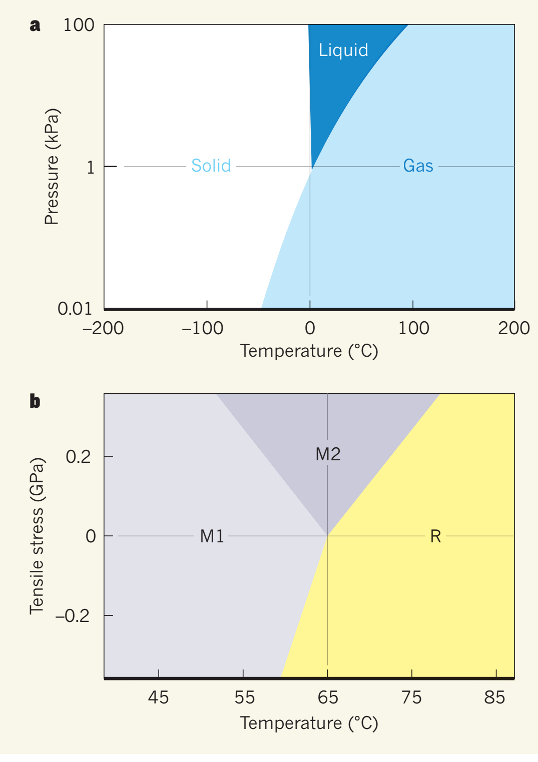


Figure 1 | Phase transitions with triple points. a, The phase diagram of water. At only one pressure and temperature can solid, liquid and gaseous water coexist in equilibrium. This triple point defines the Kelvin temperature scale. **b**, Park *et al.*[1] have mapped the phase diagram of vanadium dioxide. The triple point at zero tensile stress and the slopes of the phase boundaries greatly constrain theories that seek to understand the transitions from metal (R) to insulator (M1 or M2) in this material. Part **b** is based on Fig. 4b of the paper.

the nanobeam to ensure that the remaining suspended portions, which were no longer in contact, were stress free.

By mapping the phase diagram with high precision, the authors extracted interesting clues that constrain theoretical treatments of the phases in this system. First, it turns out that the M1, M2 and R phases can all coexist at a triple point at 65.0 °C that coincides with zero applied stress. There is no obvious reason why the M2 phase should become thermodynamically stable as soon as the tensile stress exceeds zero, as Park *et al.* observed. This fact, long obscured by lack of control over sample stresses, is something that a microscopic theory should explain. Second, the authors determined the ratio of the resistivities of the two insulating phases, a parameter that in clean crystalline material is related to the densities of states at the energy of the highest occupied electronic states and the effective masses of the electrons, quantities that should be calculable using electronic-structure methods. Finally, they determined the entropy difference per VO_2 group between the metallic phase and each insulating phase at the triple point. Although first-principles electronic-structure calculations are challenging, particularly when trying to understand effects at temperatures far above absolute zero, some future computational approach should be able to assess the relative contributions of electronic and structural degrees of freedom to these differences, further illuminating the role of electronic correlations in the transition(s).

Many other correlated systems exhibit similar phase competitions, including the manganites[4] and the rare-earth nickelates[5]. The current work shows the power of measurements that can combine micrometre- or nanometre-scale single-crystal materials, control of the stress and strain, spatial mapping of phases and *in situ* electronic transport. The importance of material quality and stresses (for example due to lattice mismatch of a film with a substrate) have long been known, and studies of bulk samples under applied and chemical pressure have been revealing in some correlated systems[6]. With the synthesis of novel structures and experimental methods such as those described here, the prospects are bright for new insights into these incredibly rich, complex materials. ■

Douglas Natelson *is in the Department of Physics and Astronomy, Rice University, Houston, Texas 77005, USA.*
e-mail: natelson@rice.edu

1. Park, J. H. *et al. Nature* **500,** 431–434 (2013).
2. Terrien, J. *Metrologia* **4,** 41 (1968).
3. Wu, J. Q. *et al. Nano Lett.* **6,** 2313–2317 (2006).
4. Dagotto, E., Hotta, T. & Moreo, A. *Phys. Rep.* **344,** 1–153 (2001).
5. Torrance, J. B., Lacorre, P., Nazzal, A. I., Ansaldo, E. J. & Niedermayer, C. *Phys. Rev. B* **45,** 8209–8212 (1992).
6. Kimber, S. A. J. *et al. Nature Mater.* **8,** 471–475 (2009).

METABOLISM

Sweet enticements to move

The formation of new blood vessels from pre-existing ones is a carefully orchestrated dance. A study reveals that the metabolism of sugar by glycolysis contributes to its regulation.

CHOLSOON JANG & ZOLTAN ARANY

The breakdown of fuel by metabolism is the engine that sustains all cellular activities. But can metabolism also steer and control cellular processes? Writing in *Cell*, De Bock *et al.*[1] suggest that the answer is yes, at least in the context of glucose metabolism and angiogenesis — the formation of new blood vessels*.

Glycolysis is the cellular process by which glucose is converted into pyruvate. A cell then makes a choice: it can convert pyruvate to lactate, which exits the cell, for a net yield of 2 ATP molecules (the currency of cellular energy transfer) or, in the presence of oxygen, the pyruvate can enter cellular organelles called mitochondria and become fully oxidized, producing a net yield of more than 30 ATP molecules. One would not expect any oxygenated cell to opt out of this mitochondrial bonanza, but some do, in a phenomenon first noted[2] in cancer cells by Otto Warburg in 1956. Cancer cells probably make this choice because intermediate molecules formed during glycolysis support the synthesis of macromolecules needed for cellular replication[3]. But do any non-cancerous or even quiescent cells also display the Warburg effect? Endothelial cells, which line blood vessels throughout the body and mediate angiogenesis, do[3], but until now little was known about how metabolism

*This article and the paper under discussion[1] were published online on 14 August 2013.

50 Years Ago

Outline of Human Genetics. By Prof. L. S. Penrose — Throughout, Prof. Penrose deals with just those points which are of general interest and particularly topics about which people ignorant of genetics are always asking, for example, Is natural selection still operating in spite of civilization and medical advances? ... In "Commentary" he explains in more detail how common chromosomal abnormalities, such as those causing mongolism and intersexes, are produced; mentions theories dealing with the possibility of inherited cancer; touches on pharmacogenetics; and outlines the vast amount of genetic variability which is being shown up by the complicated polymorphisms of the blood proteins. Finally, he makes the very good point that while geneticists are continually worrying about the quality of the human race we shall have doubled our numbers in the next 50 years and that birth control is far more important than the fruitless task of planning the superman.
From *Nature* 24 August 1963.

100 Years Ago

An exhibit illustrating the damage caused to biscuits sent out in soldered tins for the use of the troops in South Africa—especially during the Boer war—Gibraltar, Malta, Ceylon, &c., has just been placed in the central hall of the British Museum (Natural History), where it will be kept open about a month. The larvae of certain minute moths and beetles were the active agents; and it appears that since these cannot, in all probability, withstand the high temperature to which the biscuits are subjected in baking, the eggs must be laid by the moths during the period when the biscuits are being cooled before tinning.
From *Nature* 21 August 1913.

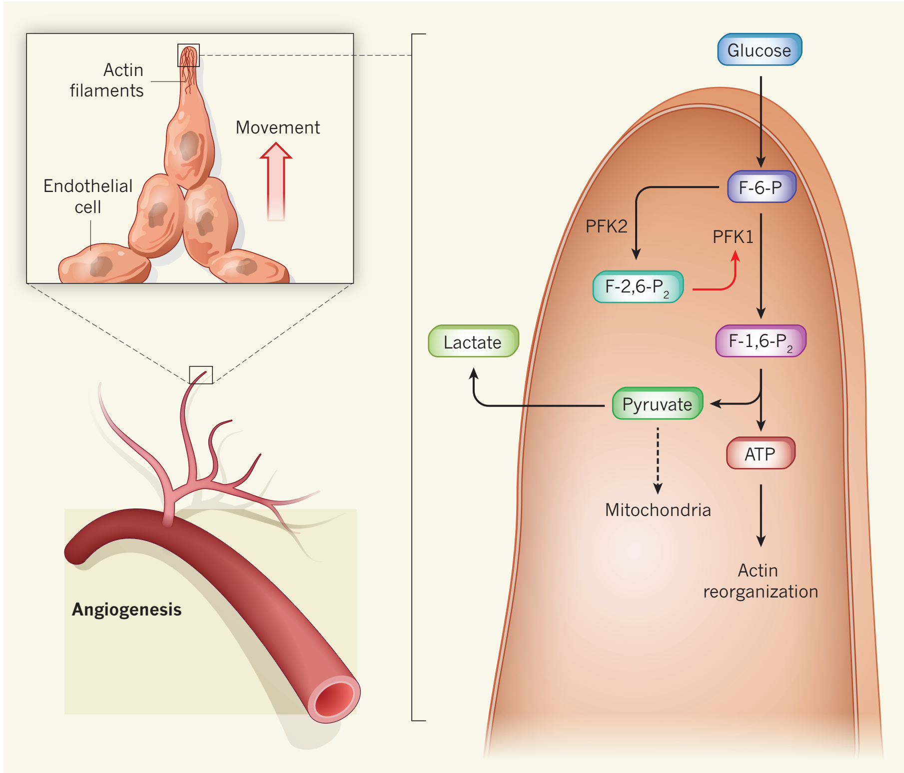


Figure 1 | Glycolysis regulates angiogenesis. The formation of new blood vessels involves the outward movement of endothelial cells from the lining of existing blood vessels, a process that relies on the rapid reorganization of actin-protein filaments in cellular structures called filopodia and lamellipodia (not shown). The energy for this (in the form of ATP) is provided by the breakdown of glucose, but endothelial cells are unusual in that the pyruvate produced by glycolysis is converted to lactate, rather than being channelled into mitochondria for further oxidation, as occurs in most cells. De Bock *et al.*[1] show that both angiogenesis and glycolysis are accelerated by the activity of the enzyme PFK2 in endothelial-cell lamellipodia and filopodia. PFK2 converts the glycolytic intermediate fructose-6-phosphate (F-6-P) into fructose-2,6-bisphosphate (F-2,6-P$_2$), which, in turn, enhances the activity of the glycolytic enzyme PFK1, thereby accelerating glycolysis at these sites. Pyruvate then leaves the cell as lactate, probably because filopodia and lamellipodia are too small to accommodate mitochondria.

affects their function.

De Bock *et al.* began their study by confirming a previous report[4] that endothelial cells are highly glycolytic but perform little pyruvate oxidation. The authors then asked the interesting question: could modulation of glycolytic activity have an effect on angiogenesis? To assess this, they altered the amount of phosphofructokinase 2 (PFK2) in endothelial cells. PFK2 is a glycolysis-regulating enzyme that was discovered only in the 1980s, long after all key enzymes of the glycolytic pathway were thought to be known[5]. The related enzyme PFK1, identified decades before PFK2, catalyses the crucial committing step of glycolysis: the conversion of fructose-6-phosphate to fructose-1,6-bisphosphate. PFK2, by contrast, converts fructose-6-phosphate to fructose-2,6-bisphosphate, which is a potent allosteric activator of PFK1 (ref. 6). Activation of PFK2 thus drastically accelerates glycolytic flux through PFK1.

De Bock *et al.* show that reducing PFK2 levels in endothelial cells not only lowers glycolytic flux, as expected, but also impairs angiogenesis, by reducing the ability of the cells to form tip cells, migrate and form blood-vessel 'sprouts'. Conversely, and importantly, increasing PFK2 levels has the opposite effect: angiogenesis is increased. The authors also show that PFK2 lies downstream of VEGF and Notch, two proteins that are dominant determinants of endothelial-cell characteristics during angiogenesis.

How does PFK2 achieve these effects? Perhaps most interestingly, the authors demonstrate that PFK2 localizes to structures at the margins of endothelial cells called lamellipodia and filopodia. These cellular projections, which contain meshes and filaments of the protein actin, mediate endothelial-cell movement and sprout formation during angiogenesis (Fig. 1). PFK2 activity at this site probably coincides with the cellular position of large complexes of glycolytic enzymes, known as metabolons, which facilitate the channelling of metabolic products from one enzyme to the next[7]. Thus, it seems that PFK2 alters angiogenic capacity by altering glycolytic flux at the site of primary cell motion.

The study is important for several reasons. The findings imply that glucose metabolism can 'steer' the angiogenic process, in addition to simply being its 'engine'. This unveils

glucose metabolism as a potential target for pro-angiogenic therapies (such as in patients with inadequate blood supply to the heart or limbs) or anti-angiogenic therapies (for example, to tackle tumours). Metabolic enzymes make good drug targets, so this is an exciting possibility. The study also provides an additional explanation for why endothelial cells perform glycolysis rather than oxidative breakdown of glucose: rapid local generation of ATP can occur in glycolytic metabolons located in the lamellipodia and filopodia, which are too small to accommodate mitochondria and are often found at angiogenic fronts where oxygen is scant.

Like all seminal work, this study generates several questions. Does modulation of glycolytic flux in ways other than through PFK2 also affect angiogenic sprouting? Could non-enzymatic properties of PFK2 contribute to the observed phenomena? Such behaviour has been seen for pyruvate kinase, another key enzyme in glycolysis that was recently found[8] to be present in the cell nucleus and associated with transcription factors that drive gene expression. Does PFK2 modulate the activities of Rac, Akt and eNOS — key enzymes that regulate endothelial-cell motility — and, if so, how? How do Notch and VEGF signal to PFK2? Does glycolysis regulate migration of other cell types, such as smooth-muscle cells or macrophages, or even cancer cells? And is the pro-angiogenic activity of PFK2 altered when glucose homeostasis is perturbed, such as in diabetes?

These questions aside, De Bock and colleagues' study deepens our understanding of why some cells choose to forego the lucrative use of mitochondria to break down their glucose, even when, as is the case for endothelial cells, the cells are not highly replicative. The authors' findings also introduce a new concept in endothelial biology: that metabolic decisions can regulate the endothelial phenotype, as well as vice versa. It turns out that, much like children, endothelial cells that gorge on sugar become hyperactive. ■

Cholsoon Jang *and* **Zoltan Arany** *are at the Cardiovascular Institute and the Center for Vascular Biology Research, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts 02215, USA.*
e-mail: zarany@bidmc.harvard.edu

1. De Bock, K. *et al. Cell* **154,** 651–663 (2013).
2. Warburg, O. *Science* **124,** 269–270 (1956).
3. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. *Science* **324,** 1029–1033 (2009).
4. Dobrina, A. & Rossi, F. *Biochim. Biophys. Acta* **762,** 295–301 (1983).
5. Van Schaftingen, E. & Hers, H. G. *Biochem. Biophys. Res. Commun.* **101,** 1078–1084 (1981).
6. Van Schaftingen, E., Hue, L. & Hers, H. G. *Biochem. J.* **192,** 897–901 (1980).
7. Srere, P. A. *Annu. Rev. Biochem.* **56,** 89–124 (1987).
8. Luo, W. *et al. Cell* **145,** 732–744 (2011).

ECOLOGY

Abundant equals nested

How ecological network structures are influenced by species coexistence, community stability and perturbations is a topic of debate. It seems that one overlooked correlate of nested structures is species abundances. SEE LETTER P.449

COLIN FONTAINE

Understanding the mechanisms that shape biodiversity is one of the main goals of ecology. Network approaches, which integrate species and the interactions among them into a single framework, have proved enlightening, revealing distinct 'architectural' patterns that are strongly associated with particular ecological interactions. For mutualistic networks — those in which the interactions benefit both partners, such as between a plant and its pollinator, or a fish and a cleaner fish — the pervasive pattern seems to be a nested one, whereby specialist species (which have few partners) interact with a subset of the many partners of more generalist species. The origin and implications of nestedness remain strongly debated. On page 449 of this issue, Suweis *et al.*[1] bring an innovative and intriguing contribution to this topic by demonstrating strong relationships among species abundances, nested architecture and community stability.

Nestedness is a pattern characterized by several features (Fig. 1), including a skewed distribution of the number of interacting partners per species, with many specialist species and few extremely generalist species. Nestedness also implies asymmetric specialization, such that specialist species tend to interact with generalist ones. Finally, the generalist species in the nested network form a single, highly connected core, making the networks very cohesive.

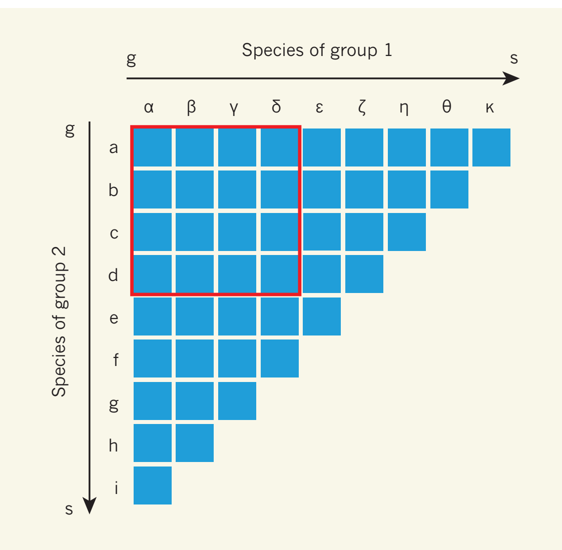


Figure 1 | A nested network. The interactions between two groups of mutualist species often assume a nested structure, in which specialist species (s), which have few partners, interact with a subset of the many partners of more generalist species (g). Here, the intersection of a row and column is blue if the species interact. Nested networks have certain characteristics, such as a continuum from highly generalist to specialist species, a core of highly connected species (red box) and a tendency for specialist species to interact with generalists (for example, the specialist i interacts with the generalist α). Suweis *et al.*[1] show that the abundances of species in an mutualistic network are positively related to the nestedness of the network.

Three main hypotheses have been proposed to explain the biology behind this seemingly highly organized structure. One is that nestedness is 'neutral', meaning that all interactions between individuals are equally likely. Species abundances in many communities are well described by a log-normal distribution, with many rare species and a few common ones. Under this hypothesis, differences in species abundance result in differences in interactions at the species level: abundant species are expected to interact more frequently and with more species than rare species, and rare species tend to interact with abundant species rather than with other rare species. However, the empirical correlation between species abundances and species generalism is not easy to interpret[2]. Do species become generalists because they are more abundant, or are they more abundant because they are generalists and therefore can access more resources?

The second hypothesis suggests that nestedness affects ecological dynamics, particularly species coexistence and community stability. A simple argument supporting this hypothesis is that it is much safer for specialist species to interact with generalist species than with other specialists, because generalist species are expected to have less-fluctuating population dynamics and so to be more reliable partners. Such constraints

on community persistence or stability could therefore be a driving force shaping interaction networks. However, no consensus on this topic has been reached among several investigations[3–6] in recent years of the links between network nestedness and community dynamics in mutualistic species.

According to the third hypothesis, nested architecture may be shaped by the (co-)evolutionary dynamics of species interacting within a community. There are many examples of interspecies interactions affecting the fitness of individuals, and of the evolution of species traits controlling the identity of potential interaction partners. Closely related species in mutualistic interaction networks tend to have similar interacting partners, which emphasizes the idea that evolutionary history has an impact on the structure of mutualistic networks[7]. But, so far, no precise evolutionary process has been directly related to a nested structure.

Suweis *et al.* have drawn these three hypotheses together by demonstrating a two-step relationship between species abundances in a community and the nestedness of the interaction network that depicts that community. Using analytical and simulation approaches, the authors first show that, under stationary conditions that have a constant number and strength of mutualistic interactions, 'interaction swaps' (an exchange of interactions between two species couples) that lead to an increase in the abundance of the species also increase the total abundance of the community. Second, the researchers demonstrate that total community abundance is positively related to the nestedness of the network. This connection opens up fascinating perspectives.

To demonstrate the implications of their findings, the authors show that, under the condition that exchanges result in increased species abundance, iterative swapping ultimately converts random networks, with randomly distributed interactions among species, into nested networks. The interpretation of this is that any process that maximizes species abundance through changes in interspecies interactions will lead to a nested network. The question thus becomes, what biological process could select for higher population size? Selection at the population level involves group-selection processes such as hard selection[8,9]. More work is needed to unravel the microevolutionary processes that affect network architecture, but this line of research seems promising.

Suweis and colleagues further demonstrate that the population size of the rarest species in the community is positively related to community resilience — the speed at which community dynamics return to equilibrium after a small perturbation. These results fuel the current debate about the relationship between network architecture and community stability[3–6,10,11] by introducing the distribution of species abundance as a key element. Again, however, the processes through which the abundance of the rarest species relates to community resilience remain to be identified. They may involve the rarest species directly, or may emerge from other mechanisms affecting both the rarest species and community resilience.

Last but not least, the relationship found by Suweis *et al.* between network nestedness and total community abundance goes both ways. Abundance is correlated with biomass, which is one of the main variables used in studies of biodiversity and ecosystem functioning, so the two-way relationship provides a bridge between the authors' results and the rich literature on these topics. We already know that the structure of food webs, for example, can affect the relationship between biodiversity and ecosystem function[12]. But little is known about the impact of mutualistic networks on the functioning of ecological communities. Like all exciting pieces of research, Suweis and colleagues' work raises more questions than it answers. ■

Colin Fontaine *is a CNRS researcher at the Muséum National d'Histoire Naturelle, Paris, UMR7204 CESCO, France.*
e-mail: cfontaine@mnhn.fr

1. Suweis, S., Simini, F., Banavar, J. R. & Maritan, A. *Nature* **500,** 449–452 (2013).
2. Santamaría, L. & Rodríguez-Gironés, M. A. *PLoS Biol.* **5,** e31 (2007).
3. Bastolla, U. *et al. Nature* **458,** 1018–1020 (2009).
4. James, A., Pitchford, J. W. & Plank, M. J. *Nature* **487,** 227–230 (2012).
5. Thébault, E. & Fontaine, C. *Science* **329,** 853–856 (2010).
6. Allesina, S. & Tang, S. *Nature* **483,** 205–208 (2012).
7. Rezende, E. L., Lavabre, J. E., Guimarães, P. R., Jordano, P. & Bascompte, J. *Nature* **448,** 925–928 (2007).
8. Saccheri, I. & Hanski, I. *Trends Ecol. Evol.* **21,** 341–347 (2006).
9. Goodnight, C. J. *Phil. Trans. R. Soc. B* **366,** 1401–1409 (2011).
10. Saavedra, S. & Stouffer, D. B. *Nature* **500,** E1–E2 (2013).
11. James, A., Pitchford, J. W. & Plank, M. J. *Nature* **500,** E2–E3 (2013).
12. Thébault, E. & Loreau, M. *Proc. Natl Acad. Sci. USA* **100,** 14949–14954 (2003).

EVOLUTIONARY BIOLOGY

A gut feeling for isolation

The far-reaching effects of the relationship between an animal and its resident gut microorganisms are becoming ever clearer. New findings suggest it can even create barriers that keep species separate.

GREGORY D. D. HURST & CHRIS D. JIGGINS

The process of speciation, whereby one lineage splits into two independent gene pools, has at its heart the evolution of barriers to gene flow that maintain differences when populations are in contact. Gene flow can be reduced in many ways, including failure to mate, sperm–egg incompatibility, and sterility or inviability of hybrids. Writing in *Science*, Brucker and Bordenstein[1] describe a novel source of reproductive isolation: the influence of resident gut microorganisms on hybrid survival.

The concept of microbial involvement in reproductive isolation is not new[2]. In the 1990s, it was recognized[3] that the very low survival rates of hybrid offspring from two closely related wasp species was influenced by the presence and strain of *Wolbachia* bacteria in the parents. More recently, it was demonstrated that environmentally induced changes in the composition of the gut microbiota could affect mate preference in *Drosophila* fruitflies[4]. Brucker and Bordenstein's work likewise examined the role of gut microbiota in reproductive isolation, but focused on the death of hybrid larvae rather than mate preference, and studied a situation in which the pool of environmental microbes was constant.

Their study organisms were parasitic wasps of the genus *Nasonia*, which lay their eggs in the pupae of flesh and filth flies (Fig. 1). The fly host represents both a source of nutrition and an environmental pool of microbes, and the authors had previously established[5] that different *Nasonia* species acquire distinct communities of resident gut microorganisms (their 'gut microbiomes') from this common microbial pool. This differentiation of microbiomes was linked to the hosts' phylogeny: the microbiomes of individuals from the closely related species *Nasonia giraulti* and *Nasonia longicornis* were more similar than that of a more distantly related species, *Nasonia vitripennis*. Brucker and Bordenstein hypothesized that this differentiation creates a setting in which dysfunctional interactions could arise between hybrids and their gut microbiota.

To test this idea, the authors examined the male offspring formed by crosses between *N. vitripennis* and *N. giraulti*, most of which

die during larval development. They noted that dying larvae were melanized, a characteristic of microbial pathology. Furthermore, the gut microbiota of the hybrid larvae were dominated by *Proteus mirabilis* bacteria, in contrast to those of the parental species, which are dominated by *Providencia* species. By manipulating the exposure of the larvae to bacteria, the researchers established that the presence of gut microorganisms was necessary for hybrid pathology and death: the hybrid larvae had near-normal fitness when they were reared on a bacteria-free diet, but their viability declined when *Proteus* and *Providencia* bacteria were introduced to the culture medium together, and when *Providencia* were introduced alone. Hybrid lethality was also reinstated when *Escherichia coli* bacteria, which are not typically found in the guts of these wasps, were introduced to the bacteria-free medium. Intriguingly, the authors also found that several gene-variant combinations in the host genome that have previously been associated with hybrid inviability were present in normal inheritance ratios in larvae reared on a bacteria-free diet.

ROBERT M. BRUCKER

Figure 1 | Laid down for life. This scanning electron micrograph shows a *Nasonia vitripennis* wasp laying eggs into the pupa of a *Sarcophaga bullata* flesh fly. The eggs (blue) hatch about 24 hours after being laid, and the larvae (purple) remain under the outer casing of the pupa for about nine days, using the fly as a nutrient source. Brucker and Bordenstein[1] show that bacteria acquired during feeding, which are normally symbiotic with the wasp, can kill hybrid wasp larvae.

These findings demonstrate that hybrid inviability can be associated with a perturbed host–microbiome interaction. Indeed, there are reasons to believe that the involvement of symbiotic microorganisms in reproductive isolation may be common. First, resident microbiota affect both organismal development and function[6], and this is likely to be the case in all species with a gut. Second, divergences in animal–microbiome interactions between lineages are widely observed[2,7], and lineage divergence is the core requirement for hybrid dysfunction. Dietary shifts, for instance, alter the microbes to which a species is exposed and those that flourish in its gut[7,8]. The different characteristics of these organisms and the differing services they provide to the host may create selective pressures for modification of host systems. Furthermore, evolutionary changes in host systems can modify the constitution of the microbiome[5,7]. In sum, there is good reason to believe that the host–microbiome interface will diverge over time, from both sides, such that hybridization creates interface combinations that may malfunction.

Brucker and Bordenstein previously extended[2] the concept of genetic interactions leading to hybrid dysfunction (called Bateson–Dobzhansky–Muller incompatibilities) to include potential incompatibilities between the genomes of the host and its microbiota. But the bacteria-reintroduction experiments in the current study suggest that, although the inviability of *Nasonia* hybrids results from disharmony of gut bacteria, this is not particularly dependent on the identity (nor, therefore, the specific genetics) of the microbes. Rather, it seems that hybrid *Nasonia* mishandle what is possibly the most fundamental and universal biotic interaction: the development and regulation of their microbiome. This breakdown echoes other cases in which biotic interactions are important in creating hybrid inviability. In *Heliconius* butterflies, for example, related species diverge to have distinct warning colouration patterns that reduce predation by birds, but hybrids between the species have a pattern that is not recognized, resulting in increased predation and a form of 'extrinsic' hybrid inviability[9].

The work of Brucker and Bordenstein provokes several questions. For example, how many cases of hybrid inviability derive from a misfunctioning interface between the host and its microbiota? Can dietary shifts drive evolutionary divergence at the host–microbiome interface, and thereby contribute to the evolution of reproductive isolation? Perhaps most exciting is the idea that there may be interactions with specific microbiome components in hybrids. If this were the case, the microbiome would expand the network of possible interactions within an organism, and potentially accelerate the rate at which incompatibility evolves. Our gut feeling is clear: the evolutionary biology of the intimate and complex interactions between animals and microbes will be a hot topic in the years to come. ■

Gregory D. D. Hurst *is at the Institute of Integrative Biology, University of Liverpool, Liverpool L697ZB, UK.* **Chris D. Jiggins** *is in the Department of Zoology, University of Cambridge, Cambridge CB23EJ, UK.*
e-mails: g.hurst@liverpool.ac.uk; c.jiggins@zoo.cam.ac.uk

1. Brucker, R. M. & Bordenstein, S. R. *Science* **341,** 667–669 (2013).
2. Brucker, R. M. & Bordenstein, S. R. *Trends Ecol. Evol.* **27,** 443–451 (2012).
3. Breeuwer, J. A. J. & Werren, J. H. *Nature* **346,** 558–560 (1990).
4. Sharon, G. *et al. Proc. Natl Acad. Sci. USA* **107,** 20051–20056 (2010).
5. Brucker, R. M. & Bordenstein, S. R. *Evolution* **66,** 349–362 (2012).
6. Shin, S. C. *et al. Science* **334,** 670–674 (2011).
7. Ley, R. E. *et al. Science* **320,** 1647–1651 (2008).
8. Chandler, J. A., Lang, J. M., Bhatnagar, S., Eisen, J. A. & Kopp, A. *Plos Genet.* **7,** e1002272 (2011).
9. Merrill, R. M. *et al. Proc. R. Soc. B* **279,** 4907–4913 (2012).

ARTICLE

doi:10.1038/nature12477

Signatures of mutational processes in human cancer

A list of authors and their affiliations appears at the end of the paper

All cancers are caused by somatic mutations; however, understanding of the biological processes generating these mutations is limited. The catalogue of somatic mutations from a cancer genome bears the signatures of the mutational processes that have been operative. Here we analysed 4,938,362 mutations from 7,042 cancers and extracted more than 20 distinct mutational signatures. Some are present in many cancer types, notably a signature attributed to the APOBEC family of cytidine deaminases, whereas others are confined to a single cancer class. Certain signatures are associated with age of the patient at cancer diagnosis, known mutagenic exposures or defects in DNA maintenance, but many are of cryptic origin. In addition to these genome-wide mutational signatures, hypermutation localized to small genomic regions, 'kataegis', is found in many cancer types. The results reveal the diversity of mutational processes underlying the development of cancer, with potential implications for understanding of cancer aetiology, prevention and therapy.

Somatic mutations found in cancer genomes[1] may be the consequence of the intrinsic slight infidelity of the DNA replication machinery, exogenous or endogenous mutagen exposures, enzymatic modification of DNA, or defective DNA repair. In some cancer types, a substantial proportion of somatic mutations are known to be generated by exposures, for example, tobacco smoking in lung cancers and ultraviolet light in skin cancers[2], or by abnormalities of DNA maintenance, for example, defective DNA mismatch repair in some colorectal cancers[3]. However, our understanding of the mutational processes that cause somatic mutations in most cancer classes is remarkably limited.

Different mutational processes often generate different combinations of mutation types, termed 'signatures'. Until recently, mutational signatures in human cancer have been explored through a small number of frequently mutated cancer genes, notably *TP53* (ref. 4). Although informative, these studies have limitations. To generate a mutational signature, a single mutation from each cancer sample is entered into a mutation set aggregated from several cases of a particular cancer type. A signature that contributes the large majority of somatic mutations in the tumour class is accurately reported. However, if multiple mutational processes are operative, a jumbled composite signature is generated. Furthermore, because such studies are based on 'driver' mutations[1], signatures of selection are superimposed on the signatures of mutational processes.

Recent advances in sequencing technology have overcome past limitations of scale[1]. Thousands of somatic mutations can now be identified in a single cancer sample, offering the possibility of deciphering mutational signatures even when several mutational processes are

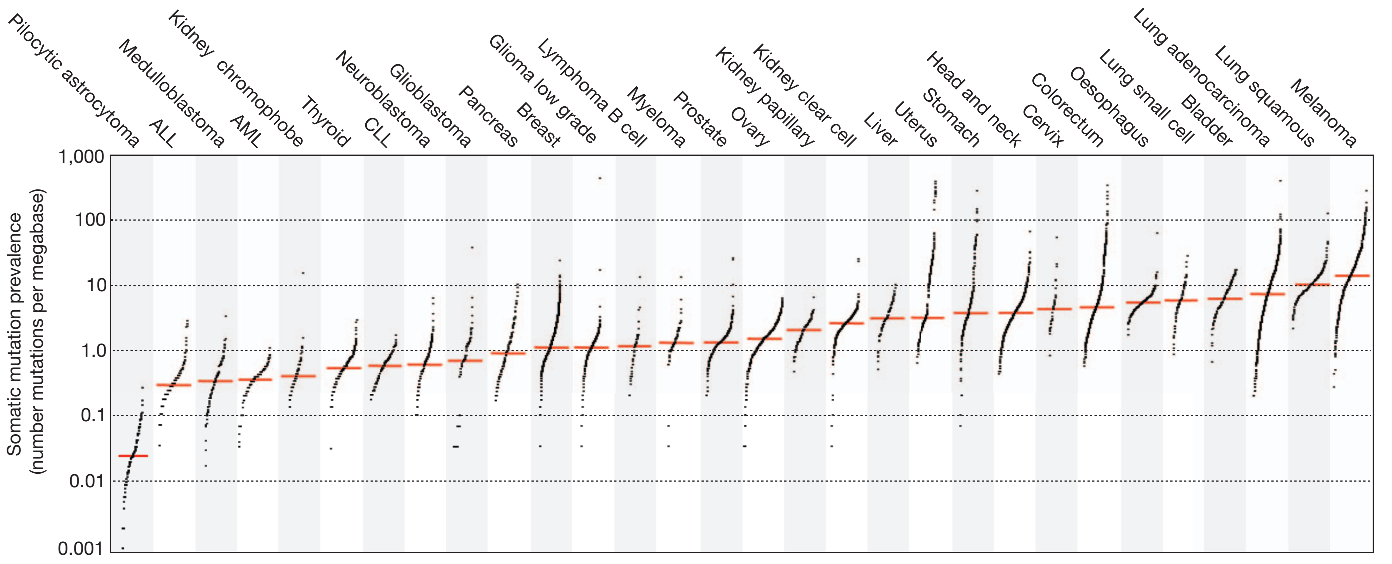


Figure 1 | The prevalence of somatic mutations across human cancer types. Every dot represents a sample whereas the red horizontal lines are the median numbers of mutations in the respective cancer types. The vertical axis (log scaled) shows the number of mutations per megabase whereas the different cancer types are ordered on the horizontal axis based on their median numbers of somatic mutations. We thank G. Getz and colleagues for the design of this figure[26]. ALL, acute lymphoblastic leukaemia; AML, acute myeloid leukaemia; CLL, chronic lymphocytic leukaemia.

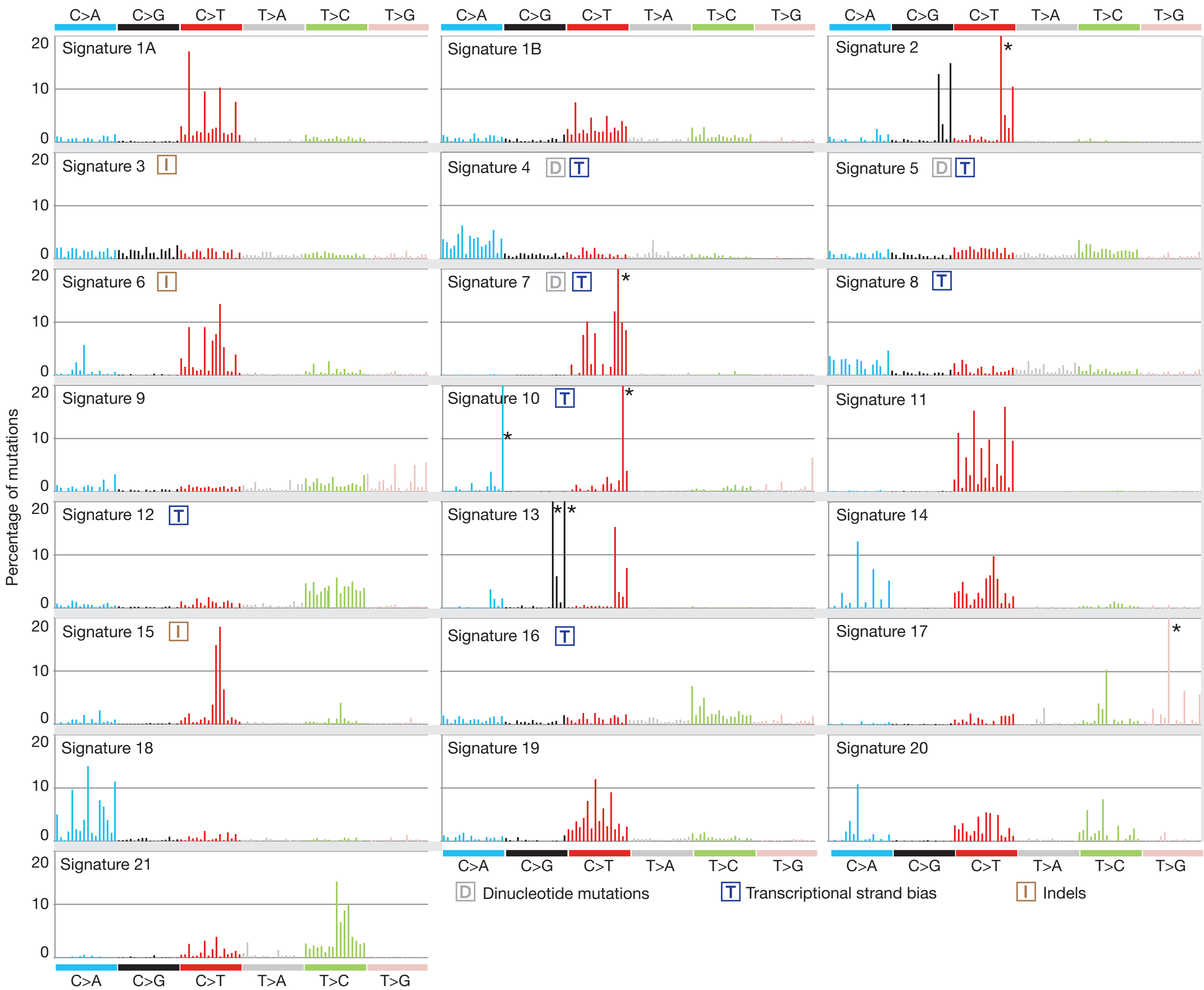


Figure 2 | Validated mutational signatures found in human cancer. Each signature is displayed according to the 96 substitution classification defined by the substitution class and sequence context immediately 3′ and 5′ to the mutated base. The probability bars for the six types of substitutions are displayed in different colours. The mutation types are on the horizontal axes, whereas vertical axes depict the percentage of mutations attributed to a specific mutation type. All mutational signatures are displayed on the basis of the trinucleotide frequency of the human genome. A higher resolution of each panel is found respectively in Supplementary Figs 2–23. Asterisk indicates mutation type exceeding 20%.

operative. Moreover, because most mutations in cancer genomes are 'passengers'[1] they do not bear strong imprints of selection.

We recently developed an algorithm to extract mutational signatures from catalogues of somatic mutations and applied it to 21 breast cancer whole-genome sequences[5,6]. Novel and known signatures were revealed, with the contribution of each signature to each cancer sample and the timing of its activity estimated[6,7]. Further studies have demonstrated that the approach can also be applied, albeit with less power, to mutational catalogues from sequences of all coding exons (exomes)[5]. Global sequencing initiatives are now yielding catalogues of somatic mutations from thousands of cancers[8]. We have therefore applied this method to survey the repertoire of mutational signatures and processes operating across the spectrum of human neoplasia.

Mutational catalogues

We compiled 4,938,362 somatic substitutions and small insertions/deletions (indels) from the mutational catalogues of 7,042 primary cancers of 30 different classes (507 from whole genome and 6,535 from exome sequences) (Supplementary Fig. 1). In all cases, normal DNA from the same individuals had been sequenced to establish the somatic origin of variants.

The prevalence of somatic mutations was highly variable between and within cancer classes, ranging from about 0.001 per megabase (Mb) to more than 400 per Mb (Fig. 1). Certain childhood cancers carried fewest mutations whereas cancers related to chronic mutagenic exposures such as lung (tobacco smoking) and malignant melanoma (exposure to ultraviolet light) exhibited the highest prevalence. This variation in mutation prevalence is attributable to differences between cancers in the duration of the cellular lineage between the fertilized egg and the sequenced cancer cell and/or to differences in somatic mutation rates during the whole or parts of that cellular lineage[1].

The landscape of mutational signatures

In principle, all classes of mutation (such as substitutions, indels, rearrangements) and any accessory mutation characteristic, for example, the sequence context of the mutation or the transcriptional strand on which it occurs, can be incorporated into the set of features by which a mutational signature is defined. In the first instance, we extracted mutational

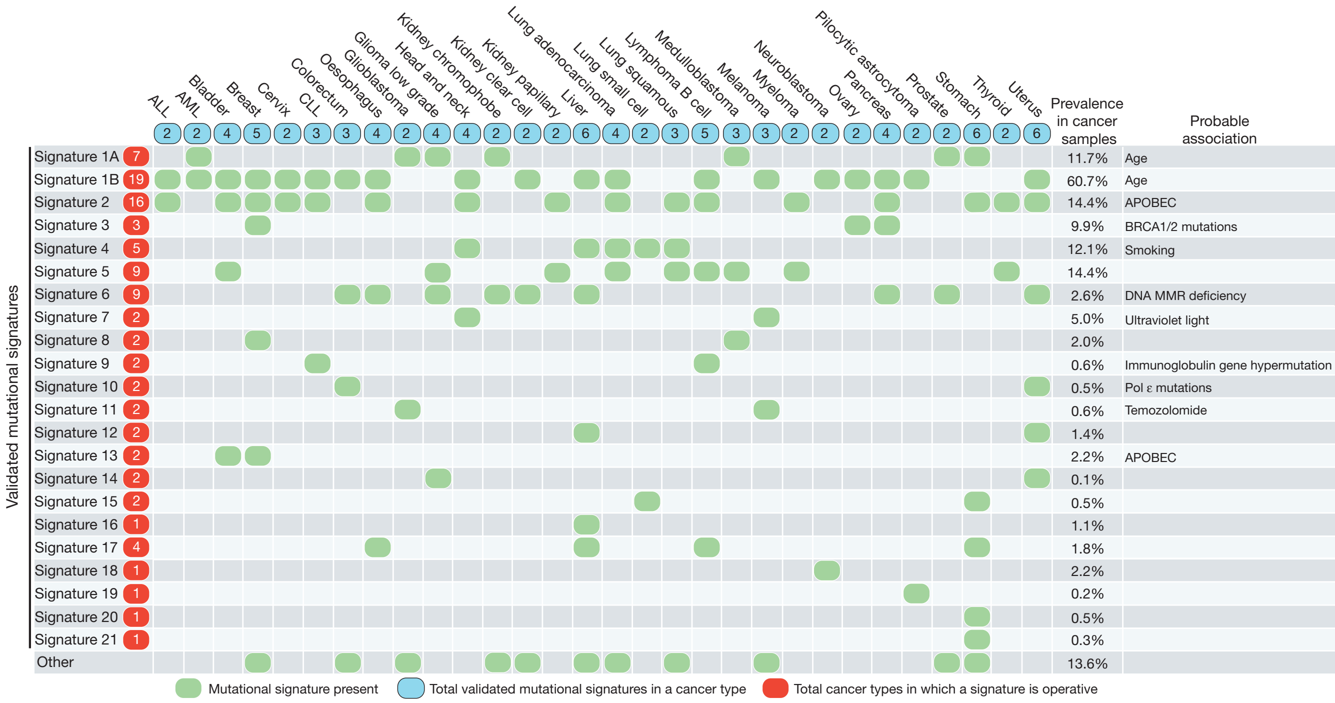


Figure 3 | The presence of mutational signatures across human cancer types. Cancer types are ordered alphabetically as columns whereas mutational signatures are displayed as rows. 'Other' indicates mutational signatures for which we were not able to perform validation or for which validation failed (Supplementary Figs 24–28). Prevalence in cancer samples indicates the percentage of samples from our data set of 7,042 cancers in which the signature contributed significant number of somatic mutations. For most signatures, significant number of mutations in a sample is defined as more than 100 substitutions or more than 25% of all mutations in that sample. MMR, mismatch repair.

signatures using base substitutions and additionally included information on the sequence context of each mutation. Because there are six classes of base substitution—C>A, C>G, C>T, T>A, T>C, T>G (all substitutions are referred to by the pyrimidine of the mutated Watson–Crick base pair)—and as we incorporated information on the bases immediately 5′ and 3′ to each mutated base, there are 96 possible mutations in this classification. This 96 substitution classification is particularly useful for distinguishing mutational signatures that cause the same substitutions but in different sequence contexts.

Applying this approach to the 30 cancer types revealed 21 distinct validated mutational signatures (Supplementary Table 1 and Supplementary Figs 2–28). These show substantial diversity (Fig. 2 and Supplementary Figs 2–23). There are signatures characterized by prominence of only one or two of the 96 possible substitution mutations, indicating remarkable specificity of mutation type and sequence context (signature 10). By contrast, others exhibit a more-or-less equal representation of all 96 mutations (signature 3). There are signatures characterized predominantly by C>T (signatures 1A/B, 6, 7, 11, 15, 19), C>A (4, 8, 18), T>C (5, 12, 16, 21) and T>G mutations (9, 17), with others showing distinctive combinations of mutation classes (2, 13, 14).

Signatures 1A and 1B were observed in 25 out of 30 cancer classes (Fig. 3). Both are characterized by prominence of C>T substitutions at NpCpG trinucleotides. Because they are almost mutually exclusive among tumour types they probably represent the same underlying process, with signature 1B representing less efficient separation from other signatures in some cancer types. Signature 1A/B is probably related to the relatively elevated rate of spontaneous deamination of 5-methyl-cytosine which results in C>T transitions and which predominantly occurs at NpCpG trinucleotides[9]. This mutational process operates in the germ line, where it has resulted in substantial depletion of NpCpG sequences, and in normal somatic cells[10].

Signature 2 is characterized primarily by C>T and C>G mutations at TpCpN trinucleotides and was found in 16 out of 30 cancer types (Fig. 3). On the basis of similarities in mutation type and sequence context we previously proposed that signature 2 is due to over activity of members of the APOBEC family of cytidine deaminases, which convert cytidine to uracil, coupled to activity of the base excision repair and DNA replication machineries[6,11].

In most cancer classes at least two mutational signatures were observed, with a maximum of six in cancers of the liver, uterus and stomach. Although these differences may, in part, be attributable to differences in the power to extract signatures, it seems likely that some cancers have a more complex repertoire of mutational processes than others.

Most individual cancer genomes exhibit more than one mutational signature and many different combinations of signatures were observed (Fig. 4 and Supplementary Figs 29–88). The patterns of contribution to individual cancer samples vary markedly between signatures. Signature 1A/B contributes relatively similar numbers of mutations to most cancer cases whereas other signatures contribute overwhelming numbers of mutations to some cancer samples but very few to others of the same cancer class, for example, signatures 2, 3, 4, 6, 7, 9, 10, 11, 13 (Fig. 4).

Mutational signatures and age of cancer diagnosis

We examined each cancer type for correlations between age of diagnosis and the number of mutations attributable to each signature in each sample. Signature 1A/B exhibited strong positive correlations with age in the majority of cancer types of childhood and adulthood (Supplementary Table 2). No other mutational signature showed a consistent correlation with age of diagnosis.

The mutations in a cancer genome may be acquired at any stage in the cellular lineage from the fertilized egg to the sequenced cancer cell. The correlation with age of diagnosis is consistent with the hypothesis that a substantial proportion of signature 1A/B mutations in cancer genomes have been acquired over the lifetime of the cancer patient, at a relatively constant rate that is similar in different people, probably in normal somatic tissues. The absence of consistent correlation of all

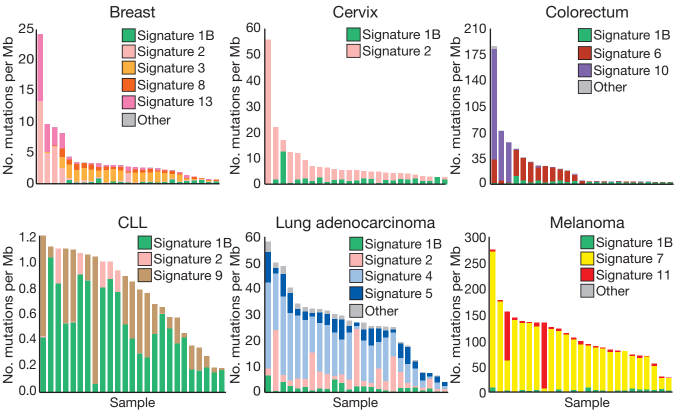


Figure 4 | The contributions of mutational signatures to individual cancers of selected cancer types. Each bar represents a typical selected sample from the respective cancer type and the vertical axis denotes the number of mutations per megabase. Contributions across all cancer samples could be found in Supplementary Figs 29–58. Summary of the total contributions for all operative mutational processes in a cancer type can be found in Supplementary Figs 59–88. 'Other' indicates mutational signatures for which we were not able to perform validation or for which validation failed (Supplementary Figs 24–28).

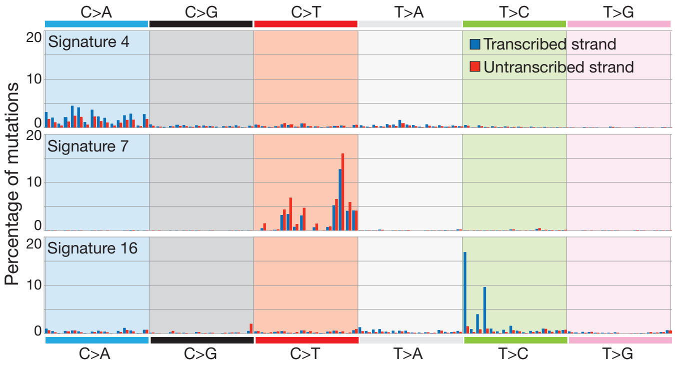


Figure 5 | Selected mutational signatures with strong transcriptional strand bias. Mutations are shown according to the 192 mutation classification incorporating the substitution type, the sequence context immediately 5′ and 3′ to the mutated base and whether the mutated pyrimidine is on the transcribed or untranscribed strand. The mutation types are displayed on the horizontal axis, whereas the vertical axis depicts the percentage of mutations attributed to a specific mutation type. A higher resolution version of all mutational signatures with strong transcriptional strand bias is found respectively in Supplementary Figs 89–95.

other signatures with age suggests that mutations associated with these have been generated at different rates in different people, possibly as a consequence of differing carcinogen exposures or after neoplastic change has been initiated.

Mutational signatures with transcriptional strand bias

The efficiency of DNA damage and DNA maintenance processes can differ between the transcribed and untranscribed strands of genes. The most well known cause of this phenomenon is transcription-coupled nucleotide excision repair (NER) that operates predominantly on the transcribed strand of genes and is recruited by RNA polymerase II when it encounters bulky DNA helix-distorting lesions[12].

We re-extracted substitution mutational signatures incorporating the transcriptional strand on which each mutation has taken place. Because a mutation in a transcribed genomic region may be either on the transcribed or the untranscribed strand, this generates a classification with 192 mutation subclasses.

Several signatures showed substantial differences in mutation prevalence between transcribed and untranscribed strands (known as transcriptional strand bias) (Fig. 5 and Supplementary Figs 89–95). For example, signature 4 shows transcriptional strand bias for C>A mutations (Fig. 5). Signature 4 is observed in lung adeno, squamous and small cell carcinomas, head and neck squamous, and liver cancers (Fig. 3), most of which are known to be caused by tobacco smoking. Therefore, signature 4 is probably an imprint of the bulky DNA adducts generated by polycyclic hydrocarbons found in tobacco smoke and their removal by transcription-coupled NER[13]. The higher prevalence of C>A mutations on transcribed compared to untranscribed strands is consistent with the propensity of many tobacco carcinogens to form adducts on guanine.

Similarly, signature 7, mainly found in malignant melanoma, shows a higher prevalence of C>T mutations on the untranscribed compared to the transcribed strands consistent with the formation, through ultraviolet exposure, of pyrimidine dimers and other lesions which are known to be repaired by transcription-coupled NER[14].

Beyond these known examples of DNA damage processed by transcription-coupled NER, other signatures show strong transcriptional strand bias (5, 8, 10, 12, 16). Notably, signature 16, which is characterized by T>C mutations at ApTpA, ApTpG and ApTpT trinucleotides and is observed in hepatocellular carcinomas, shows the strongest transcriptional strand bias of any signature, with T>C mutations occurring almost exclusively on the transcribed strand (Fig. 5). Similarly, signature 12, which features T>C mutations at NpTpN trinucleotides, also found in hepatocellular carcinomas, shows strong transcriptional strand bias with more T>C mutations on the transcribed than untranscribed strands (Supplementary Fig. 94). On the assumption that the transcriptional strand biases in signatures 12 and 16 are introduced by transcription-coupled NER, these currently unexplained signatures may be the result of bulky DNA helix-distorting adducts on adenine. However, there is no previous basis for invoking transcription-coupled NER in the genesis of these signatures and other causes of transcriptional strand bias may exist.

Mutational signatures with insertions and deletions

We re-extracted the mutational signatures including, in addition to the 96 substitution types, two further classes of mutation: indels at short nucleotide repeats and indels with overlapping microhomology at breakpoint junctions. Three of the 21 base substitution signatures associated with large numbers of indels. Signature 6, which is characterized predominantly by C>T at NpCpG mutations, but is distinct from signature 1A/B, contributes very large numbers of substitutions and small indels (mostly of 1 bp) at nucleotide repeats to subsets of colorectal, uterine, liver, kidney, prostate, oesophageal and pancreatic cancers. This pattern of indels, often termed 'microsatellite instability', is characteristic of cancers with defective DNA mismatch repair[15]. Consistent with this explanation, the presence of signature 6 was strongly associated with the inactivation of DNA mismatch repair genes in colorectal cancer ($P = 3.3 \times 10^{-5}$).

Signature 15 also contributes very large numbers of substitutions and small indels at nucleotide repeats but, compared to signature 6, exhibits greater prominence of C>T at GpCpN trinucleotides. Signature 15 was found in several samples of lung and stomach cancer and its origin is currently unknown.

By contrast, substantial numbers of larger deletions (up to 50 bp) with overlapping microhomology at breakpoint junctions were found in breast, ovarian and pancreatic cancer cases with major contributions from signature 3. A subset of cancer cases of these three classes is known to be due to inactivating mutations in *BRCA1* and *BRCA2*, and the presence of signature 3 was strongly associated with *BRCA1* and *BRCA2* mutations within the individual cancer types ($P = 1.6 \times 10^{-8}$ for breast cancer and $P = 0.02$ for pancreatic cancer)[6]. Indeed, almost all cases with *BRCA1* and *BRCA2* mutations showed a large contribution from signature 3. However, some cases with a substantial contribution from signature 3 did not have *BRCA1* and *BRCA2* mutations,

indicating that other mechanisms of *BRCA1* and *BRCA2* inactivation or abnormalities of other genes may also generate it.

BRCA1 and BRCA2 are implicated in homologous-recombination-based DNA double-strand break repair[16]. Abrogation of their functions results in non-homologous end-joining mechanisms, which can use microhomology at rearrangement junctions to rejoin double-strand breaks, taking over DNA double-strand break repair. The results show that, in addition to the genomic structural instability conferred by defective double-strand break repair, a base substitution mutational signature is associated with BRCA1 and BRCA2 deficiency.

Associating cancer aetiology and mutational signatures

Each mutational signature is the imprint left on the cancer genome by a mutational process that may include one or more DNA damage and/or DNA maintenance mechanisms, with the latter either functioning normally or abnormally. Here we consider likely mechanisms or underlying causes by comparing signatures with mutation patterns of known causation in the scientific literature or by associating them with epidemiological and biological features of particular cancer types.

Signature 1A/B is probably due to the endogenous mutational process present in most normal and neoplastic cells that is initiated by deamination of 5-methyl-cytosine[9]. Other signatures are probably attributable to exogenous mutagenic exposures. Signature 7 is observed in malignant melanoma and squamous carcinoma of the head and neck and has the known features of ultraviolet-light-induced mutations. Signature 4 is found in cancers associated with tobacco smoking (Fig. 3) and has the mutational features associated with tobacco carcinogens[13]. The causal relationship between tobacco smoking and signature 4 is supported by a strong positive association between smoking history and the contributions of signature 4 to individual cancers ($P = 1.1 \times 10^{-7}$, Supplementary Figs 44–46, 74–76 and 96).

Cigarette smoke contains over 60 carcinogens[13] and it is possible that this complex mixture may initiate other mutational processes. Signatures 1A/B, 2 and 5 were also found in lung adenocarcinoma. Signature 5, but not signatures 1A/B and 2, also showed a positive correlation between smoking history and mutation contribution ($P = 8.0 \times 10^{-3}$, Supplementary Fig. 96). Thus, in lung cancer, signature 5, which is characterized predominantly by C>T and T>C mutations, may also be due to tobacco carcinogens. However, it is also present in nine other cancer types, most of which are not strongly associated with tobacco consumption, and therefore its aetiology overall is unclear (Fig. 3).

Some anticancer drugs are mutagens[17]. Signature 11 is found in malignant melanomas and glioblastoma multiforme pretreated with the alkylating agent temozolomide ($P = 4.0 \times 10^{-3}$) and has mutational features very similar to those previously reported in experimental studies of alkylating agents[18].

Abnormalities in DNA maintenance may also be responsible for mutational signatures, and the roles of defective DNA mismatch repair (signature 6) and defective homologous-recombination-based DNA double-strand break repair (signature 3) have been discussed above. Other signatures may result from abnormal activity of enzymes that modify DNA or of error-prone polymerases. Signatures 2 and 13 have been attributed to the AID/APOBEC family of cytidine deaminases[6]. On the basis of similarities in the sequence context of cytosine mutations caused by APOBEC enzymes in experimental systems, a role for APOBEC1, APOBEC3A and/or APOBEC3B in human cancer seems more likely than for other members of the family[19–21]. However, the reason for the extreme activation of this mutational process in some cancers is unknown. Because APOBEC activation constitutes part of the innate immune response to viruses and retrotransposons[22] it may be that these mutational signatures represent collateral damage on the human genome from a response originally directed at retrotransposing DNA elements or exogenous viruses. Confirmation of this hypothesis would establish an important new mechanism for initiation of human carcinogenesis.

Signature 9, observed in chronic lymphocytic leukaemia and malignant B-cell lymphomas, is characterized by T>G transversions at Ap<u>T</u>pN and Tp<u>T</u>pN trinucleotides, and is restricted to cancers that have undergone somatic immunoglobulin gene hypermutation (IGHV-mutated) associated with AID ($P = 2.5 \times 10^{-4}$ in chronic lymphoid leukaemia (CLL)). Signature 9 does not, however, have the known mutational features of AID[20], and has been proposed to be due to polymerase η, an error-prone polymerase involved in processing AID-induced cytidine deamination[11,23]. Similarly, signature 10, which generates huge numbers of mutations in subsets of colorectal and uterine cancer, has been previously associated with altered activity of the error-prone polymerase Pol ε consequent on mutations in the gene[24,25].

Many mutational signatures do not, however, have an established or proposed underlying mutational process or aetiology. Some, for example signatures 8, 12 and 16, show strong transcriptional strand bias (Fig. 5) and possibly reflect the involvement of transcription-coupled nucleotide excision repair acting on bulky DNA adducts due to exogenous carcinogens. Others, for example signatures 14, 15 and 21, show overwhelming activity in a small number of cancer cases (Supplementary Figs 38, 45 and 56, respectively) and are perhaps more likely to be due to currently uncharacterized defects in DNA maintenance.

Localized hypermutation

Foci of localized substitution hypermutation, termed kataegis after the Greek for thunderstorm, were recently described in breast cancer[6]. Kataegis is characterized by clusters of C>T and/or C>G mutations which are substantially enriched at Tp<u>C</u>pN trinucleotides and on the same DNA strand. Foci of kataegis include from a few to several thousand mutations and are often found in the vicinity of genomic rearrangements. The genomic regions affected are different in different cancers. On the basis of the substitution types and sequence context of kataegis substitutions, an underlying role for APOBEC family enzymes was proposed for kataegis as well as for signatures 2 and 13 (ref. 6).

The 507 whole-cancer genome mutation catalogues were searched for clusters of mutations. Cancers of breast (67 of 119), pancreas (11 of 15), lung (20 of 24), liver (15 of 88), medulloblastomas (2 of 100), CLL (15 of 28), B-cell lymphomas (21 of 24) and acute lymphoblastic leukaemia (1 of 1) showed occasional (<10), small (<20 mutations) foci of kataegis, whereas acute myeloid leukaemia (0 of 7) and pilocytic astrocytoma (0 of 101) did not. Subsets of breast (7), lung (6) and haematological cancers (3) showed numerous (>10) kataegic foci and two breast and one pancreatic cancer showed major foci of kataegis (>50 mutations) (Fig. 6 and Supplementary Figs 97 and 98).

Kataegic foci are often associated with genomic rearrangements (Supplementary Fig. 98). In yeast, introduction of a DNA double-strand break greatly increases the likelihood of kataegis in its vicinity, indicating a role for such breaks in initiating the process[20]. However, even in cancer cases with kataegis, most rearrangements do not exhibit nearby kataegis, indicating that a double-strand break is not sufficient.

In neoplasms of B-lymphocyte origin, including CLL and many lymphomas, mutation clusters recurrently occurred at immunoglobulin loci. In these cancers the mutation characteristics were different (Supplementary Fig. 98), bearing the hallmarks of somatic hypermutation associated with AID, which is operative during the generation of immunological diversity[20].

Discussion

The diversity and complexity of somatic mutational processes underlying carcinogenesis in human beings is now being revealed through mutational patterns buried within cancer genomes. It is likely that more mutational signatures will be extracted, together with more precise definition of their features, as the number of whole-genome sequenced cancers increases and analytical methods are further refined.

The mechanistic basis of some signatures is, at least partially, understood but for many it remains speculative or unknown. Elucidating the

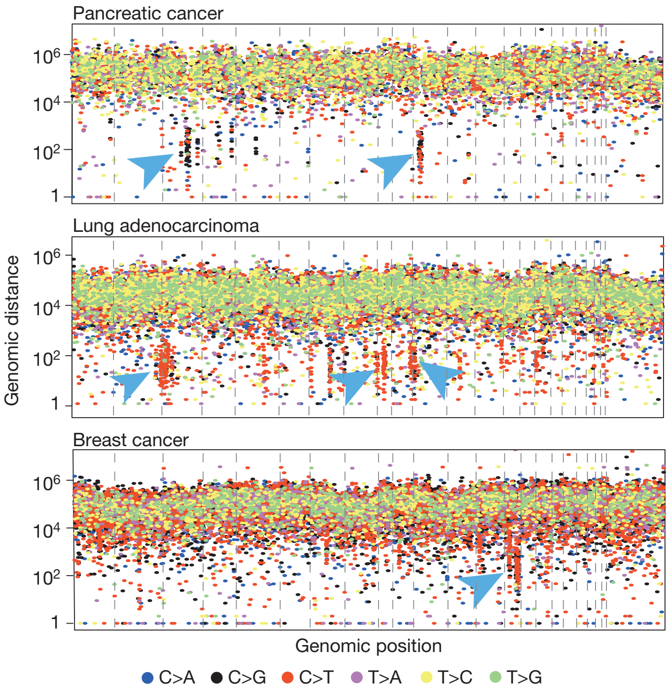


Figure 6 | Kataegis in three cancers. Each of these 'rainfall' plots represents an individual cancer sample in which each dot represents a single somatic mutation ordered on the horizontal axis according to its position in the human genome. The vertical axis denotes the genomic distance of each mutation from the previous mutation. Arrowheads indicate clusters of mutations in kataegis.

underlying mutational processes will depend upon two major streams of investigation. First, compilation of mutational signatures from model systems exposed to known mutagens or perturbations of the DNA maintenance machinery and comparison with those found in human cancers. Second, correlation of the contributions of mutational signatures with other biological characteristics of each cancer through diverse approaches ranging from molecular profiling to epidemiology. Collectively, these studies will advance our understanding of cancer aetiology with potential implications for prevention and treatment.

METHODS SUMMARY

Mutational catalogues were stringently filtered and our previously developed computational framework[5,6] was used to extract mutational signatures from them. The computational framework for deciphering mutational signatures and all mutational catalogues are freely available for download from http://www.mathworks.com/matlabcentral/fileexchange/38724, whereas the complete set of somatic mutations is available from ftp://ftp.sanger.ac.uk/pub/cancer/AlexandrovEtAl. All presented mutational signatures were validated. Kataegis was detected using an algorithm based on piecewise constant fitting.

Full Methods and any associated references are available in the online version of the paper.

Received 24 March; accepted 19 July 2013.
Published online 14 August 2013.

1. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458,** 719–724 (2009).
2. Pfeifer, G. P. Environmental exposures and mutational patterns of cancer genomes. *Genome Med.* **2,** 54 (2010).
3. Peña-Diaz, J. *et al.* Noncanonical mismatch repair as a source of genomic instability in human cells. *Mol. Cell* **47,** 669–680 (2012).
4. Olivier, M., Hollstein, M. & Hainaut, P. TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb. Perspect. Biol.* **2,** a001008 (2010).
5. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3,** 246–259 (2013).
6. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149,** 979–993 (2012).
7. Nik-Zainal, S. *et al.* The life history of 21 breast cancers. *Cell* **149,** 994–1007 (2012).
8. Hudson, T. J. *et al.* International network of cancer genome projects. *Nature* **464,** 993–998 (2010).
9. Pfeifer, G. P. Mutagenesis at methylated CpG sequences. *Curr. Top. Microbiol. Immunol.* **301,** 259–281 (2006).
10. Welch, J. S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150,** 264–278 (2012).
11. Di Noia, J. M. & Neuberger, M. S. Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* **76,** 1–22 (2007).
12. Hanawalt, P. C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nature Rev. Mol. Cell Biol.* **9,** 958–970 (2008).
13. Pfeifer, G. P. *et al.* Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* **21,** 7435–7451 (2002).
14. Pfeifer, G. P., You, Y. H. & Besaratinia, A. Mutations induced by ultraviolet light. *Mutat. Res.* **571,** 19–31 (2005).
15. Boland, C. R. & Goel, A. Microsatellite instability in colorectal cancer. *Gastroenterology* **138,** 2073–2087 (2010).
16. Thompson, L. H. Recognition, signaling, and repair of DNA double-strand breaks produced by ionizing radiation in mammalian cells: the molecular choreography. *Mutat. Res.* **751,** 158–246 (2012).
17. Hunter, C. *et al.* A hypermutation phenotype and somatic MSH6 mutations in recurrent human malignant gliomas after alkylator chemotherapy. *Cancer Res.* **66,** 3987–3991 (2006).
18. Tomita-Mitchell, A. *et al.* Mismatch repair deficient human cells: spontaneous and MNNG-induced mutational spectra in the HPRT gene. *Mutat. Res.* **450,** 125–138 (2000).
19. Taylor, B. J. M. *et al.* DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *eLife* e00534 (2013).
20. Burns, M. B. *et al.* APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494,** 366–370 (2013).
21. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell* **10,** 1247–1253 (2002).
22. Koito, A. & Ikeda, T. Intrinsic immunity against retrotransposons by APOBEC cytidine deaminases. *Front. Microbiol.* **4,** 28 (2013).
23. Puente, X. S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475,** 101–105 (2011).
24. The Cancer Genome Atlas Network.. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487,** 330–337 (2012).
25. Cancer Genome Atlas Research. Integrated genomic characterization of endometrial carcinoma. *Nature* **497,** 67–73 (2013).
26. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank the Wellcome Trust for support (grant reference 098051) together with many other funding bodies and individuals (Supplementary Note 1).

Author Contributions L.B.A., S.N.-Z. and M.R.S. conceptualized the study and analysed the mutational signatures and kataegis data. L.B.A. performed data curation, data filtering and mutational signature extraction. S.N.-Z. and D.C.W. performed kataegis identification. S.N.-Z. performed visual validation. A.P.B., K.R., J.W.T. and D.J. provided bioinformatics support for mutational signature and kataegis analysis. S.A.J.R.A., S.Be., A.V.B., G.R.B., N.B., A.B., A.-L.B.-D., S.Bo., B.B., C.C., H.R.D., C.D., R.E., J.E.E., J.A.F., M.G., F.H., B.H., T.I., S.I., M.I., N.J., D.T.W.J., S.K., M.K., S.R.L., C.L.-O., S.M., N.C.M., H.N., P.A.N., M.P., E.P., A.P., J.V.P., X.S.P., M.R., A.L.R., J.R., P.R., M.S., T.N.S., P.N.S., Y.T., A.N.J.T., R.V.-M., M.M.v.B., L.V.V., A.V.-S., N.W., L.R.Y., J.Z.-R., P.A.F., U.M., P.L., M.M., S.M.G., R.S., E.C., T.S., S.M.P. and P.J.C. contributed samples, clinical data and scientific advice. M.R.S. and L.B.A. wrote the manuscript. M.R.S. directed the overall research.

Author Information The computational framework for deciphering mutational signatures and all mutational catalogues are freely available for download from http://www.mathworks.com/matlabcentral/fileexchange/38724, whereas the complete set of somatic mutations is available from ftp://ftp.sanger.ac.uk/pub/cancer/AlexandrovEtAl. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.R.S. (mrs@sanger.ac.uk).

Ludmil B. Alexandrov[1], Serena Nik-Zainal[1,2], David C. Wedge[1], Samuel A. J. R. Aparicio[3,4,5], Sam Behjati[1,6], Andrew V. Biankin[7,8,9,10,11], Graham R. Bignell[1], Niccolò Bolli[1,12,13], Ake Borg[14], Anne-Lise Børresen-Dale[15,16], Sandrine Boyault[17], Birgit Burkhardt[18,19], Adam P. Butler[1], Carlos Caldas[20], Helen R. Davies[1], Christine Desmedt[21], Roland Eils[22], Jórunn Erla Eyfjörd[23], John A. Foekens[24], Mel Greaves[25], Fumie Hosoda[26], Barbara Hutter[22], Tomislav Ilicic[1], Sandrine Imbeaud[27,28], Marcin Imielinsk[29], Natalie Jäger[22], David T. W. Jones[30], David Jones[1], Stian Knappskog[31,32], Marcel Kool[30], Sunil R. Lakhani[33], Carlos López-Otín[34], Sancha Martin[1], Nikhil C. Munshi[35,36], Hiromi Nakamura[26], Paul A. Northcott[30], Marina Pajic[7], Elli Papaemmanuil[1], Angelo Paradiso[37], John V. Pearson[38], Xose S. Puente[34], Keiran Raine[1], Manasa Ramakrishna[1], Andrea L. Richardson[39,40,41], Julia Richter[42], Philip Rosenstiel[43], Matthias Schlesner[22], Ton N. Schumacher[44], Paul N. Span[45], Jon W.

Teague[1], Yasushi Totoki[26], Andrew N. J. Tutt[46], Rafael Valdés-Mas[34], Marit M. van Buuren[44], Laura van 't Veer[47], Anne Vincent-Salomon[48], Nicola Waddell[38], Lucy R. Yates[1], Australian Pancreatic Cancer Genome Initiative*, ICGC Breast Cancer Consortium*, ICGC MMML-Seq Consortium*, ICGC PedBrain*, Jessica Zucman-Rossi[27,28], P. Andrew Futreal[1], Ultan McDermott[1], Peter Lichter[49], Matthew Meyerson[29,39,40], Sean M. Grimmond[38], Reiner Siebert[42], Elías Campo[50], Tatsuhiro Shibata[26], Stefan M. Pfister[30,51], Peter J. Campbell[1,12,13] & Michael R. Stratton[1]

[1]Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK. [2]Department of Medical Genetics, Box 134, Addenbrooke's Hospital NHS Trust, Hills Road, Cambridge CB2 0QQ, UK. [3]Molecular Oncology, Michael Smith Genome Sciences Centre, BC Cancer Agency, 675 West 10th Avenue, Vancouver V5Z 1L3, Canada. [4]Centre for Translational and Applied Genomics, Michael Smith Genome Sciences Centre, BC Cancer Agency, 675 West 10th Avenue, Vancouver V5Z 1L3, Canada. [5]Department of Pathology, University of British Columbia, G227-2211 Wesbrook Mall, British Columbia, Vancouver V6T 2B5, Canada. [6]Department of Paediatrics, University of Cambridge, Hills Road, Cambridge CB2 2XY, UK. [7]Wolfson Wohl Cancer Research Centre, Institute of Cancer Sciences, University of Glasgow, Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1BD, UK. [8]West of Scotland Pancreatic Unit, Glasgow Royal Infirmary, Glasgow G4 0SF, UK. [9]The Kinghorn Cancer Centre, 370 Victoria Street, Darlinghurst, and the Cancer Research Program, Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, Sydney, New South Wales 2010, Australia. [10]Department of Surgery, Bankstown Hospital, Eldridge Road, Bankstown, Sydney, New South Wales 2200, Australia. [11]South Western Sydney Clinical School, Faculty of Medicine, University of New South Wales, Liverpool, New South Wales 2170, Australia. [12]Department of Haematology, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK. [13]Department of Haematology, University of Cambridge, Cambridge CB2 2XY, UK. [14]Department of Oncology, Lund University, SE-221 85 Lund, Sweden. [15]Department of Genetics, Institute for Cancer Research, Oslo University Hospital, The Norwegian Radium Hospital, Montebello, 0310 Oslo, Norway. [16]The K.G. Jebsen Center for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, N-0310 Oslo, Norway. [17]Plateforme de Bioinformatique Synergie Lyon Cancer, Centre Léon Bérard, 28 rue Laennec, 69373 Lyon Cedex 08, France. [18]NHL-BFM Study Center and Department of Pediatric Hematology and Oncology, University Children's Hospital, 48149 Münster, Germany. [19]NHL-BFM Study Center and Department of Pediatric Hematology and Oncology, University Children's Hospital, 35392 Giessen, Germany. [20]Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge CB2 0RE, UK. [21]Breast Cancer Translational Res Lab -BCTL, Université Libre de Bruxelles—Institut Jules Bordet, Boulevard de Waterloo, 125, B-1000 Brussels, Belgium. [22]Department of Theoretical Bioinformatics (B080), German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. [23]Cancer Research Laboratory, Faculty of Medicine, Biomedical Centre, University of Iceland, 101 Reykjavik, Iceland. [24]Department of Medical Oncology, Erasmus MC Cancer Institute, 3015 CE Rotterdam, The Netherlands. [25]Department of Haemato-oncology, Institute of Cancer Research, London SM2 5NG, UK. [26]Division of Cancer Genomics, National Cancer Center Research Institute, Chuo-ku, Tokyo 104-0045, Japan. [27]INSERM, UMR-674, Génomique Fonctionnelle des Tumeurs Solides, Institut Universitaire d'Hematologie (IUH), 75475 Paris, France. [28]Université Paris Descartes, Labex Immuno-oncology, Sorbonne Paris Cité, Faculté de Médecine, 75006 Paris, France. [29]The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02141, USA. [30]Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [31]Section of Oncology, Department of Clinical Science, University of Bergen, 5020 Bergen, Norway. [32]Department of Oncology, Haukeland University Hospital, 5021 Bergen, Norway. [33]The University of Queensland Centre for Clinical Research, School of Medicine and Pathology Queensland, The Royal Brisbane & Women's Hospital, Herston 4029, Brisbane, Queensland, Australia. [34]Departamento Bioquímica y Biología Molecular, IUOPA-Universidad de Oviedo, 33006 Oviedo, Spain. [35]Jerome Lipper Multiple Myeloma Disease Center, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02215, USA. [36]Boston Veterans Administration Healthcare System, West Roxbury, Massachusetts 02132, USA. [37]Clinical Experimental Oncology Laboratory, National Cancer Institute, Via Amendola, 209, 70126 Bari, Italy. [38]Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, The University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia. [39]Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, Massachusetts 02215, USA. [40]Harvard Medical School, Boston, Massachusetts 02115, USA. [41]Department of Pathology, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA. [42]Institute of Human Genetics, Christian-Albrechts-University, 24118 Kiel, Germany. [43]Institute of Clinical Molecular Biology, Christian-Albrechts-University, 24118 Kiel, Germany. [44]Division of Immunology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. [45]Department of Radiation Oncology and department of Laboratory Medicine, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500HB Nijmegen, The Netherlands. [46]Breakthrough Breast Cancer Research Unit, King's College London School of Medicine, London SW3 6JB, UK. [47]The Netherlands Cancer Institute, 121 Plesmanlaan, 1066 CX Amsterdam, The Netherlands. [48]Institut Curie, Departement de Pathologie, INSERM U830, 26 rue d'Ulm, 75248 Paris Cedex 05, France. [49]Division of Molecular Genetics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [50]Unidad de Hematopatología, Servicio de Anatomía Patológica, Hospital Clínic, Universitat de Barcelona, IDIBAPS, 08036 Barcelona, Spain. [51]Department of Pediatric Hematology and Oncology, 69120 Heidelberg, Germany.

*A list of authors and affiliations appears in the Supplementary Information.

METHODS

Validating mutational signatures. Validating a mutational signature requires ensuring that a large set of somatic mutations attributed to this signature is genuine in at least one sample. Validation is complicated as multiple mutational processes are usually operative in most cancer samples, and thus every individual somatic mutation can be probabilistically assigned to several mutational signatures. To overcome this limitation, we examined our data set for samples that are predominantly generated by one mutational signature (that is, more than 50% of the somatic mutations in the sample belong to an individual mutational signature) and/or for samples in which all operative mutational processes have mutually exclusive patterns of mutations (for example, a sample with mutations only from signature 1B, which is predominantly C>T substitutions, and signature 18, which is predominantly C>A substitutions). We identified the optimal available sample for every mutational signature and attempted to validate the subset of somatic mutations attributed to this signature using one of three methods (Supplementary Fig. 99): (1) validation through re-sequencing with an orthogonal sequencing technology; (2) validation through re-sequencing with the same sequencing technology (including RNA-seq, bisulphite sequencing, etc.); (3) validation through visual examination of somatic mutations by an experienced curator using a genomic browser and BAM files for both the tumour and its matched normal.

For some of the previously published samples, we used the already reported validation data. When possible, somatic mutations were validated by either re-sequencing with orthogonal technology or re-sequencing using the same sequencing technology. We resorted to visual validation only when there was no other possibility for validating a mutational signature. 22 out of the 27 originally identified mutational signatures were validated (Supplementary Table 1 and Supplementary Fig. 99). Three mutational signatures failed validation: signatures R1 to R3 (Supplementary Figs 24 to 26). We were unable to validate two mutational signatures: signatures U1 and U2 (Supplementary Figs 27 and 28), due to lack of available biological samples and access to BAM files for the samples with sufficient number of somatic mutations generated by these two mutational signatures.

Samples and curation of freely available cancer data. Informed consent was obtained from all subjects. Collection and use of patient samples were approved by the appropriate Internal Review Board of each institution. In addition to newly generated data, we curated freely available somatic mutations from three other sources: (1) the data portal of The Cancer Genome Atlas (TCGA); (2) the data portal of the International Cancer Genome Consortium (ICGC); (3) previously published data in peer-review journals, see additional references[6,23,27–59].

Filtering, estimating mutation prevalence and generating mutational catalogues. In all examined samples, normal DNA from the same individuals had been sequenced to establish the somatic origin of variants. Extensive filtering was performed to remove any residual germline mutations and technology-specific sequencing artefacts before analysing the data. Germline mutations were filtered out from the lists of reported mutations using the complete list of germline mutations from dbSNP[60], 1000 genomes project[61], NHLBI GO Exome Sequencing Project[62], and 69 Complete Genomics panel (http://www.completegenomics.com/public-data/69-Genomes/). Technology-specific sequencing artefacts were filtered out by using panels of BAM files of (unmatched) normal tissues containing more than 120 normal genomes and 500 normal exomes. Any somatic mutation present in at least three well-mapping reads in at least two normal BAM files was discarded. The remaining somatic mutations were used for generating a mutational catalogue for every sample.

Prevalence of somatic mutations was estimated on the basis of a haploid human genome after all filtering. Prevalence of somatic mutations in exomes was calculated based on the identified mutations in protein-coding genes and assuming that an average exome has 30 Mb in protein-coding genes with sufficient coverage. Prevalence of somatic mutations in whole genomes was calculated based on all identified mutations and assuming that an average whole genome has 2.8 gigabases with sufficient coverage.

The immediate 5′ and 3′ sequence context was extracted using the ENSEMBL Core programing interfaces for human genome build GRCh37. Curated somatic mutations that originally mapped to an older version of the human genome were remapped using UCSC's freely available lift genome annotations tool (any somatic mutations with ambiguous or missing mappings were discarded). Dinucleotide substitutions were identified when two substitutions were present in consecutive bases on the same chromosome (sequence context was ignored). The immediate 5′ and 3′ sequence content of all indels was examined and the ones present at mono/polynucleotide repeats or microhomologies were included in the analysed mutational catalogues as their respective types. Strand bias catalogues were derived for each sample using only substitutions identified in the transcribed regions of well-annotated protein-coding genes. Genomic regions of bidirectional transcription were excluded from the strand bias analysis.

Deciphering signatures of mutational processes. Mutational signatures were deciphered independently for each of the 30 cancer types using our previously developed computational framework[5]. The algorithm deciphers the minimal set of mutational signatures that optimally explains the proportion of each mutation type found in each catalogue and then estimates the contribution of each signature to each catalogue. Mutational signatures were also extracted separately for genomes and exomes. Mutational signatures extracted from exomes were normalized using the observed trinucleotide frequency in the human exome to the one of the human genome. All mutational signatures were clustered using unsupervised agglomerative hierarchical clustering and a threshold was selected to identify the set of consensus mutational signatures. Mis-clustering was avoided by manual examination (and whenever necessary re-assignment) of all signatures in all clusters. 27 consensus mutational signatures were identified across the 30 cancer types. The computational framework for deciphering mutational signatures as well as the data used in this study are freely available and can be downloaded from http://www.mathworks.com/matlabcentral/fileexchange/38724, whereas the complete set of somatic mutations is available from ftp://ftp.sanger.ac.uk/pub/cancer/AlexandrovEtAl.

Factors that influence extraction of mutational signatures. Recently, using simulated and real data, we described in detail the factors that influence the extraction of mutational signatures[5]. These included the number of available samples, the mutation prevalence in samples, the number of mutations contributed by different mutational signatures, the similarity between the signatures of mutational processes operative in cancer samples, as well as the limitations of our computational approach. Here, we examined data sets with varying sizes from 30 different cancer types and we have taken great care to report only validated mutational signatures. However, our approach identified two similar patterns most likely representing the same biological process; that is, signature 1A and 1B. The reasons for this is, for some cancer types we have sufficient numbers of samples and/or mutations (that is, statistical power) to decipher the cleaner version (that is, signature 1A), whereas for other cancer types we do not have sufficient data and our approach extracts a version of the signature which is more contaminated by other signatures present in that cancer type (that is, signature 1B). Nevertheless, the two signatures are very similar; hence we call them 1A and 1B. Being almost mutually exclusive among cancer types (that is, finding either signature 1A or 1B in each cancer type but not usually both) is supportive of the notion that they represent the same underlying process as is the fact that signatures 1A and 1B both correlate with age and have the same overall pattern of contributions to individual cancer genomes. Indeed, in our view it is likely that if we had sufficient data, signature 1B would disappear and the algorithm would extract only signature 1A.

Displaying mutational signatures. Mutational signatures are displayed using a 96 substitution classification defined by the substitution class and the sequence context immediately 3′ and 5′ to the mutated base. Mutational signatures are displayed in the main text of the report and in Supplementary Information on the basis of the observed trinucleotide frequency of the human genome; that is, representing the relative proportions of mutations generated in each signature based on the actual trinucleotide frequencies of the reference human genome. However, in Supplementary Information we also provide a visualization of mutational signatures based on an equal frequency of each trinucleotide (Supplementary Figs 2–28). The equal trinucleotide frequency representation results, in all mutational signatures, in a greater degree of prominence of C>T substitutions at Np<u>C</u>pG trinucleotides as major features compared to the plots based on the observed trinucleotides. This difference may in some cases reflect the biological reality, that is, a propensity of the particular mutational process to be more active at NpCpG trinucleotides. However, note that it may also in some cases be due to incomplete extraction by the algorithm of the signature in question from signature 1A/B, which is characterized by prominent features at Np<u>C</u>pG trinucleotides. This is likely to happen because (1) signature 1A/B is ubiquitous and (2) because even a small probability of mutations at Np<u>C</u>pG trinucleotides will generate a prominent feature because of the severe depletion of Np<u>C</u>pG trinucleotides in the reference genome. In future, with larger numbers of sequences and large numbers of whole-genome sequences it is anticipated that the latter effect will be reduced.

Approaches for associating cancer aetiology and exposures of validated mutational signatures. Generalized linear models (GLMs) were used to fit signature exposures (that is, number of mutations assigned to a signature) and age of cancer diagnoses. For each cancer type, all mutational signatures operative in it were evaluated using GLMs and the *P* values were corrected for multiple hypothesis testing using the Benjamini–Hochberg false discovery rate procedure. The resulting *P* values indicate that age strongly correlates with signature 1A/B across 15 cancer types (Supplementary Table 2). Exposure to signature 4 also correlates with age of diagnosis in kidney papillary and thyroid cancers. However, in both cancer types, we were not able to detect/extract signature 1A/B due to a low number of mutations in their samples and it is likely that signature 1A/B is

currently mixed within signature 4. Further studies involving whole-genome sequences will be needed to validate this hypothesis. Notably, in melanoma, age of diagnosis also correlates with exposure to signature 7, which we have associated with exposure to ultraviolet light.

Associations between all other aetiologies and signature exposures were performed using two-sample Kolmogorov–Smirnov tests between two sets of samples. The first set contains the signature exposures of the samples with the 'desired feature' (for example, samples that contain a hypermutation in the immunoglobulin gene) and the second set is the signature exposures of the samples without the 'desired feature' (for example, samples that do not contain a hypermutation in the immunoglobulin gene). Samples with unknown feature status (for example, not knowing the status of the immunoglobulin gene) were ignored. Kolmogorov–Smirnov tests were performed for all signatures and all examined 'features' in a cancer type. P values were corrected for multiple hypothesis testing using the Benjamini–Hochberg false discovery rate procedure and based on the performed tests in a particular cancer class.

A piecewise-constant-fitting-based algorithm for the detection of kataegis. Foci of localized hypermutation, termed kataegis, were sought in 507 whole-genome sequenced cancers. High-quality variant calls that had been previously subjected to filtering for mutational signature analysis were investigated using an algorithm developed to identify foci of kataegis.

For each sample, all mutations were ordered by chromosomal position and the intermutation distance, defined as the number of base pairs from each mutation to the next one, was calculated. Intermutation distances were then segmented using the piecewise constant fitting (PCF) method[63] to find regions of constant intermutation distance. Parameters used for PCF were $\gamma = 25$ and $k_{\min} = 2$ and were trained on the set of kataegis foci that had been manually identified, curated and validated using orthogonal sequencing platforms[6]. Putative regions of kataegis were identified as those segments containing six or more consecutive mutations with an average intermutation distance of less than or equal to 1,000 bp.

Variation in number of foci of kataegis and relationship with genome-wide mutation burden. To examine the likelihood of kataegis occurring for different mutation burdens, the expected number of kataegis events that would be observed by chance was calculated for a range of total number of mutations per cancer, n, between 1,000 and 2,000,000. The probability that any one mutation will be followed by five other mutations within a distance of 5,000 bp, thereby triggering the identification of kataegis, is given by $p = P(\mathrm{Pois}(5{,}000n/g) \geq 5)$, where g is the length of the genome, in base pairs.

Supplementary Fig. 97 shows the expected number of kataegis events identified in genomes with between 100,000 and 500,000 mutations. For cancers with up to 200,000 mutations, the expected number of kataegis events is extremely small (0.16 for a total mutation load of 200,000), making the detection of kataegic foci highly significant for each sample. Supplementary Table 3 presents all the samples in which kataegic foci were identified, the total mutation burden for each sample, the observed number of kataegic foci, and the expected number of foci.

Specificity of variants in kataegis foci. Clusters of variant calls can easily occur in regions of low sequence complexity. These are not true substitution mutations but represent systematic sequencing artefacts or mis-mapping of short reads. The quality of variant calls depends on the quality of mutation-calling by individual institutions. Additional filtering was applied to remove likely false-positive calls and then putative kataegic foci were individually curated.

1,436 kataegis foci were called by PCF, with 873 finalized as putative kataegis foci (Supplementary Table 4) involving 9,219 substitution variants. Where possible, BAM files were retrieved, inspected and substitution variants involved in kataegis foci were manually curated to remove likely false-positive calls. Where BAM files were not available to us, substitution variants were strictly excluded if called in: (1) genomic features that generate mapping errors, for example, regions of excessively high coverage due to collapsed repeat sequences in the reference genome[64]; (2) highly repetitive regions with reads consistently demonstrating low mapping qualities in 20 unrelated normal samples; (3) locations with known germline insertions/deletions within the sequencing reads reporting the mutated base.

Several features were seen in the finalized putative kataegis foci, which reinforced the conviction in the validity of these calls. Although clusters of mutations identified by the PCF method were sought in an approach unbiased by mutation type and based exclusively on intermutation distances, we find that the 873 putative foci demonstrate: first, a preponderance to C>T and C>G mutations (Supplementary Fig. 97b); second, the enrichment for a Tp$\underline{\text{C}}$ sequence context as previously described[6] (Supplementary Fig. 97b); third, processivity (where consecutive mutations within a cluster were on the same strand; that is, 6 C>T mutations in a row or 6 G>A mutations in a row; Fig. 6c); and fourth, visual curation of reads carrying these processive variants showed that the variants were usually in *cis* (that is, mutations were on the same read (Supplementary Fig. 97c) or on the read mate of other affected alleles within the insert size) with respect to each other, indicating that they had arisen on the same allele. Finally, where data were available, we found that clusters of substitution mutations within the same kataegis foci shared approximately the same variant allele fraction, indicating that they had probably arisen during a single cell cycle event.

BAM files from some samples were not accessible and therefore a proportion of substitution variants involved in kataegis foci were not visually curated. The application of the strict criteria described above and the subsequent finding of the consistency of the mutation-type, sequence context, processive nature of the mutations, with the majority in *cis* on individual sequencing reads, indicates that the vast majority of these foci are probably genuine. However, the possibility that some of the foci are not truly kataegis, particularly for the cancers which have not been validated or visually curated, remains.

Sensitivity of kataegis detection. It is acknowledged that the likelihood of detection of kataegis foci rests on the sensitivity of mutation detection. It is possible for foci to be missed because the mutations were not detected by mutation callers of the various institutions, before our analysis. This is particularly relevant for subclonal mutations bearing a low variant allele fraction or for mutations that occur on a single copy of a multi-copy locus. This is because the likelihood of mutation detection is reduced when uncorrected for copy number and for aberrant cell fraction of the tumour sample. Furthermore, our stringent post-processing criteria, particularly of samples that have not been visually curated, make it more likely that kataegis is under-represented in this analysis.

Relationship between kataegis and large-scale genomic changes. Reinforcing our previous findings[6], we found that some kataegic foci were very closely associated with rearrangements. For example, a breast cancer sample with 1,534 point mutations had only one focus of kataegis which contained 32 point mutations. The same breast cancer sample also had 25 large-scale genomic structural variations scattered throughout the genome. However, one tandem duplication coincided with this single locus of kataegis in this cancer. Notably, no other mutations or structural variations were seen for 2 Mb flanking this extraordinary event (Supplementary Fig. 97b). Another breast cancer (Fig. 6) that contained 22,454 mutations and had 292 rearrangements altogether, had nine regions of kataegis, five of which coincided with large-scale structural variations, underscoring the co-localization of kataegis foci with structural variations. This also highlights that not all foci of kataegis co-localized with structural variations and not all structural variations were associated with kataegis.

Sites of amplification represent a potential source of false variant calls. If the amplification occurred early in the evolution of a cancer, then there is an increased likelihood of substitutions accumulating randomly within the amplified genomic region. When mapped back to the reference genome, these will appear as clustered variants.

A number of features allow us to distinguish such events from 'true' kataegis. These mutations would not be expected to have features associated with kataegis, such as the mutation type, predilection for a Tp$\underline{\text{C}}$ sequence context and the processivity. Furthermore, if they have accumulated as random events in a multi-copy locus, then they would be less likely to occur in *cis* (on the same sequencing read) with respect to each other. In contrast, mutations which have occurred at the same time, during one moment of transient hypermutability in a single cell cycle event, would be expected to cluster on one copy of a multi-copy locus, to be in *cis* and to demonstrate approximately the same variant allele fraction. Finally, to achieve the level of hypermutation required to be called as a focus of kataegis (average intermutation distance of less than 1,000 bp for six consecutive mutations equivalent to ~1,000 substitutions per Mb), the degree of copy number amplification would have to be considerable.

To examine this likelihood of false calls in regions of amplification, simulations were performed assuming background mutation rates of 10 per Mb, 40 per Mb and 100 per Mb for different copy number states and for different sizes of focal amplification. The expected number of kataegic foci for these different states are provided in Supplementary Table 5. For most of the samples in which kataegis was detected (all but twenty), a 10 Mb region of amplification would require a copy number state of 36 or above to generate 1 cluster of 6 mutations with an average intermutation distance of less than 1,000 bp. For 19 of the remaining 20 samples, a 10 Mb region of amplification would require a copy number state of 10 or above. For the single cancer with a mutation rate exceeding 40 per Mb, a copy number state of 4 is required to generate a cluster of mutations. As mentioned previously, these clusters would have to be processive, be in *cis* and have roughly the same variant allele fraction to be called as a focus of kataegis.

Definition of kataegis. Kataegis has been identified via a PCF-based method as 6 or more consecutive mutations with an average intermutation distance of less than or equal to 1,000 bp. Other salient features include a preponderance for C>T and C>G mutations, a predilection for a Tp$\underline{\text{C}}$ mutation context, processivity, evidence of having arisen on the same parental allele (being in *cis*) on sequencing reads and additionally (but not necessarily) co-localization with large-scale genomic structural variation.

27. Holmfeldt, L. *et al.* The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nature Genet.* **45,** 242–252 (2013).
28. Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481,** 157–163 (2012).
29. De Keersmaecker, K. *et al.* Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. *Nature Genet.* **45,** 186–190 (2013).
30. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481,** 506–510 (2012).
31. Stephens, P. J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486,** 400–404 (2012).
32. Quesada, V. *et al.* Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature Genet.* **44,** 47–52 (2012).
33. Seshagiri, S. *et al.* Recurrent R-spondin fusions in colon cancer. *Nature* **488,** 660–664 (2012).
34. Dulak, A. M. *et al.* Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nature Genet.* **45,** 478–486 (2013).
35. Agrawal, N. *et al.* Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* **333,** 1154–1157 (2011).
36. Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333,** 1157–1160 (2011).
37. Guo, G. *et al.* Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. *Nature Genet.* **44,** 17–19 (2012).
38. Peña-Llopis, S. *et al.* BAP1 loss defines a new class of renal cell carcinoma. *Nature Genet.* **44,** 751–759 (2012).
39. Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455,** 1069–1075 (2008).
40. Seo, J. S. *et al.* The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res.* **22,** 2109–2119 (2012).
41. Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150,** 1107–1120 (2012).
42. Love, C. *et al.* The genetic landscape of mutations in Burkitt lymphoma. *Nature Genet.* **44,** 1321–1325 (2012).
43. Zhang, J. *et al.* Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. *Nature Genet.* **45,** 602–612 (2013).
44. Morin, R. D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476,** 298–303 (2011).
45. Jiao, Y. *et al.* DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science* **331,** 1199–1203 (2011).
46. Pugh, T. J. *et al.* The genetic landscape of high-risk neuroblastoma. *Nature Genet.* **45,** 279–284 (2013).
47. Jones, S. *et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330,** 228–231 (2010).
48. Wu, J. *et al.* Whole-exome sequencing of neoplastic cysts of the pancreas reveals recurrent mutations in components of ubiquitin-dependent pathways. *Proc. Natl Acad. Sci. USA* **108,** 21188–21193 (2011).
49. Sausen, M. *et al.* Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. *Nature Genet.* **45,** 12–17 (2013).
50. Berger, M. F. *et al.* The genomic complexity of primary human prostate cancer. *Nature* **470,** 214–220 (2011).
51. Grasso, C. S. *et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487,** 239–243 (2012).
52. Barbieri, C. E. *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nature Genet.* **44,** 685–689 (2012).
53. Rudin, C. M. *et al.* Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. *Nature Genet.* **44,** 1111–1116 (2012).
54. Peifer, M. *et al.* Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nature Genet.* **44,** 1104–1110 (2012).
55. Stark, M. S. *et al.* Frequent somatic mutations in MAP3K5 and MAP3K9 in metastatic melanoma identified by exome sequencing. *Nature Genet.* **44,** 165–169 (2012).
56. Berger, M. F. *et al.* Melanoma genome sequencing reveals frequent PREX2 mutations. *Nature* **485,** 502–506 (2012).
57. Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* **150,** 251–263 (2012).
58. Zang, Z. J. *et al.* Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nature Genet.* **44,** 570–574 (2012).
59. Wang, K. *et al.* Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nature Genet.* **43,** 1219–1223 (2011).
60. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29,** 308–311 (2001).
61. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
62. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493,** 216–220 (2013).
63. Baumbusch, L. O. *et al.* Comparison of the Agilent, ROMA/NimbleGen and Illumina platforms for classification of copy number alterations in human breast tumors. *BMC Genomics* **9,** 379 (2008).
64. Pickrell, J. K., Gaffney, D. J., Gilad, Y. & Pritchard, J. K. False positive peaks in ChIP-seq and other sequencing-based functional assays caused by unannotated high copy number regions. *Bioinformatics* **27,** 2144–2146 (2011).

doi:10.1038/nature12478

GLUTAMATE RECEPTOR-LIKE genes mediate leaf-to-leaf wound signalling

Seyed A. R. Mousavi[1], Adeline Chauvin[2], François Pascaud[3], Stephan Kellenberger[3] & Edward E. Farmer[1]

Wounded leaves communicate their damage status to one another through a poorly understood process of long-distance signalling. This stimulates the distal production of jasmonates, potent regulators of defence responses. Using non-invasive electrodes we mapped surface potential changes in *Arabidopsis thaliana* after wounding leaf eight and found that membrane depolarizations correlated with jasmonate signalling domains in undamaged leaves. Furthermore, current injection elicited jasmonoyl-isoleucine accumulation, resulting in a transcriptome enriched in RNAs encoding key jasmonate signalling regulators. From among 34 screened membrane protein mutant lines, mutations in several clade 3 *GLUTAMATE RECEPTOR-LIKE* genes (*GLRs 3.2*, *3.3* and *3.6*) attenuated wound-induced surface potential changes. Jasmonate-response gene expression in leaves distal to wounds was reduced in a *glr3.3 glr3.6* double mutant. This work provides a genetic basis for investigating mechanisms of long-distance wound signalling in plants and indicates that plant genes related to those important for synaptic activity in animals function in organ-to-organ wound signalling.

Unlike plants, animals rely on rapid nervous systems to escape predation. A stationary fly that perceives danger takes less than 300 ms to take off, and this process requires complex whole-body coordination[1]. Nevertheless, this escape response is too slow if the fly lands on a Venus flytrap, a plant in which electrical signals initiate rapid trap closure[2]. Whereas fast movements associated with insect capture are exceptional, slower herbivore-induced defence gene expression is widespread in plants and is coordinated between organs[3]. What, then, is the nature of the long distance signal(s) that leads to defence responses throughout much of a plant body after wounding? Among the many scenarios proposed to explain the nature of systemic wound signals in plants[4] is a role for electrical signalling[5]. However, this has not been substantiated and it is essential to identify genes that underlie this phenomenon.

Resistance to herbivores depends to a large extent on the production of potent regulatory lipids known as jasmonates[6]. Without the ability to produce or perceive these compounds, plants that normally resist attack become remarkably vulnerable to predation[7]. Both jasmonic acid (JA) and biologically active jasmonoyl-isoleucine (JA-Ile)[8] accumulate within minutes in wounds and in undamaged distal tissues[9–11]. Similarly, when feeding on *A. thaliana*, the Egyptian cotton leafworm (*Spodoptera littoralis*) stimulates jasmonate-regulated transcription in tissues several centimetres from a wound[12] and when feeding on bean, these insects provoke plasma membrane depolarizations that spread through entire wounded leaves[13]. Such plasma membrane depolarizations are common in plants[14,15] and are also produced after exposure of cells to damage-associated molecular patterns[16], including peptide danger signals[17]. Moreover, treatment of tomato cells with ionophores that cause plasma membrane depolarization stimulated the expression of jasmonate-regulated genes[18,19]. Related to this, membrane depolarizations in potato preceded increases in cytosolic Ca^{2+}, and jasmonate accumulation was reduced when these Ca^{2+} transients were blocked[20]. Here, concentrating exclusively on the jasmonate defence pathway and using non-invasive surface electrodes[21], we monitored changes in electrical activity due to ion fluxes in cell populations in wounded *Arabidopsis* leaves. We show that electrical signals activate jasmonate biosynthesis in leaves distal to wounds and identify genes involved in the propagation of these signals.

Wound-activated surface potential changes

To investigate patterns of electrical activity and gene expression in 5-week-old rosettes, individual leaves were numbered from oldest to youngest. Electrodes placed on leaf 8 at the midrib (e1 electrode position), midrib/petiole junction (e2) and on the petiole (e3) did not detect changes in electrical activity and such changes were not elicited by walking *S. littoralis* larvae (Fig. 1a–c and Extended Data Fig. 1a, b). When the larvae began to feed, wound-activated surface potential changes (WASPs) of variable amplitude, duration and complexity were observed (Supplementary Video 1 and Extended Data Fig. 1c). Because insects release chemical elicitors in addition to causing wounding[3], we investigated the effects of mechanical wounding on electrical activity. Simply touching the leaf did not generate changes in surface potential, but wounding the leaf tip resulted in strong and reproducible surface potential changes (Fig. 1b, c). When recordings were extended, they often showed periodicity (Extended Data Fig. 1d). We used three parameters to characterize these signals: latency (time from wounding to arrival at the amplitude midpoint), amplitude and duration (Fig. 1b). To gain more information on the spread of WASPs within a wounded leaf, four electrodes were placed on the leaf surface (Fig. 1a). After damage, WASPs were detected first at e1, then several seconds later at e2, and finally at e3. An electrode on the lamina also detected damage-elicited electrical activity and, in each case (Fig. 1c), the signals we measured had the same polarity as those produced after chilling, a treatment known to cause plasma membrane depolarization[22,23]. Therefore WASPs in leaf 8 were due to plasma membrane depolarization (Extended Data Fig. 1e–g).

WASP territories and speeds

Signals generated by wounding leaf tips first move towards the centre of the rosette and then disperse away from the apex into a restricted number of distal leaves to initiate distal JA accumulation and signalling[10]. To map the spatial distribution of WASPs in the rosette after wounding

[1]Department of Plant Molecular Biology, University of Lausanne, Biophore, CH-1015 Lausanne, Switzerland. [2]School of Pharmaceutical Sciences, University of Geneva, 30 quai Ernest-Ansermet, CH-1211 Geneva 4, Switzerland. [3]Department of Pharmacology and Toxicology, University of Lausanne, Rue du Bugnon 27, CH-1005 Lausanne, Switzerland.

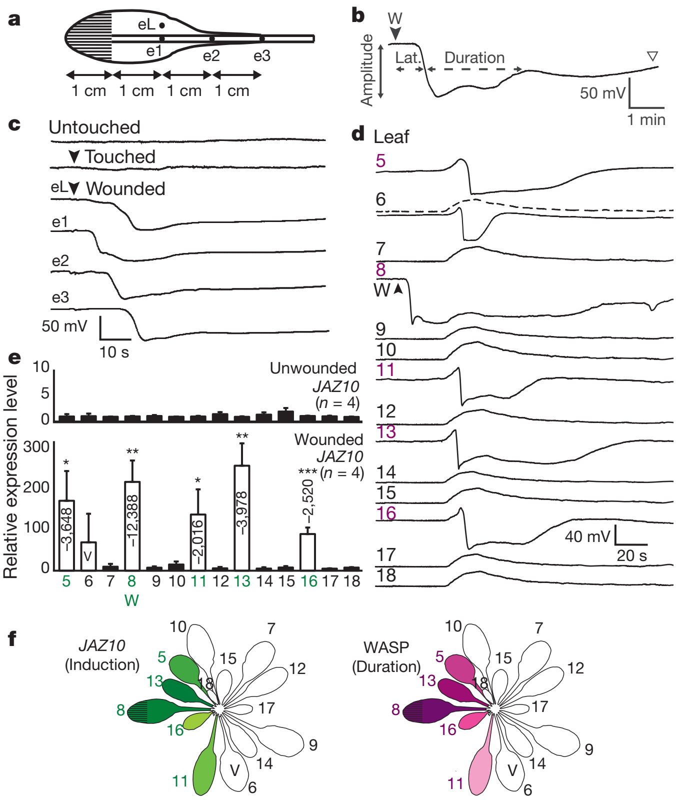


Figure 1 | WASPS and *JAZ10* expression map to identical spatial domains. **a**, Experimental design for detecting surface potential changes on leaves. Measuring electrodes e1, midrib; e2, petiole/midrib junction; e3, petiole. The lamina electrode (eL) was 3 mm from e1. The apical part of the leaf was wounded with forceps. **b**, Three distinct variables, latency (Lat.), duration and amplitude of WASPs, were analysed. The signal in the wounded leaf (leaf 8, e3) typically did not recover to baseline during recording (unfilled arrowhead). Time of wounding is indicated with a filled arrowhead. **c**, Typical surface potential changes recorded on leaf 8. Arrowheads indicate when the leaf was touched or wounded. **d**, Representative WASPs generated on distal leaves after wounding (W) leaf 8 ($n = 10$–61). Two types of surface potential change were observed on leaf 6. The solid and dashed lines show traces for 63% and 37% ($n = 19$) of events, respectively. **e**, Levels of *JAZ10* (± s.d.) transcript in unwounded leaves (upper panel) and 1 h after mechanical wounding of leaf 8 (lower panel). $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Numbers within bars show the product of amplitude and duration (mV s) for leaves that showed depolarizations. **f**, Heat maps for *JAZ10* transcript induction 1 h after wounding leaf 8 and duration of surface potential changes produced after wounding leaf 8. Only leaves that were investigated are indicated. Data for *JAZ10* levels are from **e**. WASP durations from Extended Data Fig. 2b. V, variable leaf.

leaf 8 we placed electrodes on this leaf or in the e3 position on leaves 5 through 18. The changes in amplitude observed in wounded leaf 8 were typically close to −70 mV (Extended Data Fig. 2a) and the unwounded leaves 5, 11, 13 and 16 showed similar responses (Fig. 1d, Extended Data Fig. 2b). For example, after wounding leaf 8, a WASP with a duration of 78 ± 20 s and a peak amplitude of −51 ± 9 mV was reached in leaf 13 after a latency of 66 ± 13 s ($n = 61$ plants). However, other leaves (7, 9, 10, 12, 14, 15, 17 and 18) showed small positive surface potential changes. For example, leaf 9 showed a 20 ± 5 mV change in surface potential with a latency of 54 ± 12 s ($n = 46$ plants). Most of these observations fit a developmental pattern: in adult-phase *Arabidopsis* rosettes, leaf n shares direct vascular connections to leaves $n \pm 5$ and $n \pm 8$. Thus the wounded leaf 8 is connected to leaves 13 and 16, these connections being termed parastichies[24]. Additionally, leaves 5 and 11 also showed strong negative surface potential changes after wounding leaf 8. These leaves are $n \pm 3$ relative to the wounded leaf 8 and may represent contact parastichies formed by proximal but unconnected vasculature[24]. We also recorded changes in surface potentials in the $n - 2$ leaf (leaf 6) that were similar to those in wounded leaf 8 in 63% recordings but the remaining recordings from this leaf (Fig. 1d) resembled traces from leaves such as leaf 9. We termed leaf 6 a variable leaf.

Quantitative electrophysiological data (Extended Data Fig. 2b) was then compared with transcript levels for *JASMONATE-ZIM DOMAIN 10* (*JAZ10*), a robust marker for activity of the jasmonate pathway[25]. One hour after wounding leaf 8 we detected ≥100-fold increases in *JAZ10* transcript levels in leaves 5, 8, 11, 13 and 16 (Fig. 1e). *JAZ10* transcript induction in leaf 6, like WASP production, was variable. Heat maps from quantitative data showed that *JAZ10* expression at 1 h post-wounding and WASP durations covered identical territories, spanning 137° of the rosette when variable leaf 6 ($n - 2$) was included (Fig. 1f).

We next examined the speed at which electrical signals moved within wounded leaves and from leaf to leaf. Replicated measurements indicated a range of speeds from 2.6 ± 0.6 cm min^{-1} between the wound and an electrode placed on the lamina, to up to 9 cm min^{-1} between electrodes placed on the midribs of the wounded leaf itself or at intervals along the midrib on leaf 13 (Extended Data Fig. 2c). The similar apparent velocities for surface potential changes in the midribs of wounded and distal leaves indicate that related mechanisms control electrical signalling in these leaves. However, signals from the wounded leaf seemed to slow to 5.4 ± 1.5 cm min^{-1} at the centre of the plant before accelerating again in the distal leaf, bringing the average signal speed from wounded leaf 8 to receiver leaf 13 to 5.8 ± 1.1 cm min^{-1} ($n = 13$). This overall velocity estimate is concordant with recent estimates of signal speeds based on JA accumulation in leaf 13 of *Arabidopsis* after wounding leaf 8 (ref. 26), and with self-propagating electrical activity elicited by wounding bean or barley leaves[27]. To test whether WASPs could travel from young to older leaves we wounded leaf 13 and monitored events in leaf 8. Again we observed a correlated pattern of WASP production and *JAZ10* expression (Extended Data Fig. 3). Then, to investigate whether the long-distance signals that activate jasmonate responses travel at similar speeds to WASP changes, electrodes were placed in positions e2 and e3 on leaf 8 and this leaf was wounded. When the wounded leaf 8 was severed between e2 and e3 after a signal had reached e3 we detected induced *JAZ10* expression in leaf 13. However, *JAZ10* induction was not observed if we cut the wounded leaf as the WASP arrived at e2 but before it reached e3 (Extended Data Fig. 4). We conclude that the long-distance signals that strongly activate *JAZ10* expression travel at a speed similar to electrical events elicited by wounding.

Current injection and the *Arabidopsis* transcriptome

To test for a direct link between the jasmonate pathway and electrical activity we implanted platinum (Pt) wires into the petiole of leaf 8 (Fig. 2a), injected current, and monitored the induction of surface potentials in the lamina of this leaf. By structuring the input current appropriately (40 μA for 10 s; see Methods) we were able to induce surface potential changes distal to the site of injection (Fig. 2b) without causing detectable cell damage other than that due to Pt wire implantation in the petiole, a region that was removed before analysing the lamina (Extended Data Fig. 5a–d). The signals generated in response to current injection (CI) had a mean duration of 59 ± 25 s (Extended Data Fig. 5e), similar to the durations of WASPs in leaf 13 when leaf 8 was wounded. From these data we estimated the apparent velocity of the surface potential change resulting from CI to be 6.4 ± 1.9 cm min^{-1}. This was close to an average action potential velocity of 7 cm min^{-1} that has been observed after CI in *Arabidopsis*[28]. Because CI was shown to stimulate JA accumulation in tomato[29], we measured the levels of both JA and JA-Ile after CI in leaf 8. Both compounds accumulated in response to this treatment (Fig. 2c, d).

To confirm that CI could induce jasmonate signalling the expression of two jasmonate-responsive genes, the regulatory gene *JAZ10* and *VEGETATVE STORAGE PROTEIN 2* (*VSP2*), an anti-insect

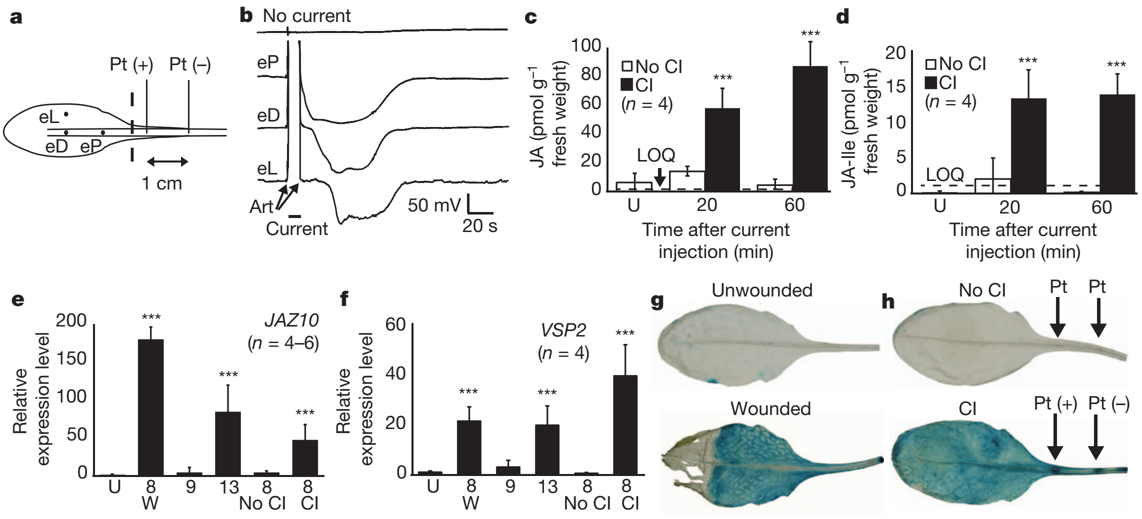


Figure 2 | Current injection (CI) induces jasmonate accumulation, gene expression and surface potential changes in leaf 8. **a**, Placement of Pt wires, proximal electrode (eP), distal electrode (eD) and laminar electrode (eL) for CI experiments. The leaf blade to the left of the dashed line was used for transcript measurements and quantification of jasmonate levels. **b**, Surface potential generation following CI (40 μA, 10 s). Art = artefacts of CI. Bar, duration of CI. **c**, Levels of JA (± s.d.) 20 min and 1 h after CI. **d**, Levels of JA-Ile (± s.d.) 20 min and 1 h after CI. LOQ, limits of quantification (dashed line). **e**, Levels of *JAZ10* transcripts (± s.d.) in leaves 1 h after wounding or CI (leaf 8 only). **f**, Levels of *VSP2* transcripts (± s.d.) in leaves 4 h after wounding or CI (leaf 8 only). U, unwounded plants. W, wounded. ***$P < 0.001$. **g**, Expression pattern of a *VSP2-GUSPlus* reporter line in leaf 8 from an unwounded plant, and 4 h after wounding. **h**, *VSP2-GUSPlus* activity in leaf 8 of control plants (no CI) and 4 h after CI.

defence gene[30], was monitored. Transcripts for both genes were upregulated in response to CI (Fig. 2e, f). Furthermore, the WASPs detected on wild-type plants were indistinguishable from those on wounded mutant plants that lacked the ability to synthesize jasmonates (Extended Data Fig. 6a–c). Therefore, the mechanism that produces WASPs is upstream or independent of jasmonate synthesis. When current was injected into the *coronatine insensitive 1-1* (*coi1-1*)[31], a mutant lacking the functional jasmonate receptor, we recorded potentials similar to those in the current-injected wild type (Extended Data Fig. 6d). However, we were unable to induce *JAZ10* expression in these plants (Extended Data Fig. 6e). The canonical jasmonate signal pathway[6] is therefore required for expression of *JAZ10* after wounding and CI. Consistent with the detection of WASPs over entire leaf surfaces, plants expressing a wound-inducible *VSP2* reporter (Fig. 2g) also responded to CI throughout the lamina (Fig. 2h).

To find out if other jasmonate-regulated genes were activated by CI we performed whole-transcriptome analysis starting with current-injected leaf 8. This revealed that 313 genes were >twofold upregulated (Fig. 3 and Supplementary Table 1). We then generated a second data set from leaf 13 one hour after wounding leaf 8. Finally, these results were compared to data produced independently from wounded 18-day-old plants[32]. This comparison showed that 94% of the CI-upregulated genes were also upregulated in leaf 13 of plants wounded on leaf 8. 70% of CI-induced transcripts were upregulated in both leaf wounding data sets. Strikingly, among these were 9 of the 12 *Arabidopsis JAZ* genes (Extended Data Fig. 7a). These genes are critical regulators of jasmonate signalling[4,6,25].

It is known that more transcripts are upregulated than are downregulated in response to wounding[12,25,32]. Consistent with this, the levels of only 66 transcripts decreased in response to CI. Of these transcripts, 47% were also downregulated in one or both wounding experiments (Extended Data Fig. 7b). Thirteen genes were downregulated in all three treatments (Extended Data Fig. 7c). Clearly, CI does not affect the expression of all wound-regulated genes and this is consistent with the fact that several signal mechanisms operate to control gene expression in damaged plants[4,25,33,34]. One such pathway depends on NADPH oxidase D (*RBOHD*) to transmit reactive oxygen species (ROS)-dependent long-distance signals at speeds similar to the WASP velocities we describe[34]. Although inhibitors of this ROS propagation pathway[34] did not abolish WASPs (Extended Data Fig. 8a–c), one of them, diphenyleneiodonium (DPI), reduced WASP duration in leaf 13 by a factor of 2. However, DPI did not inhibit distal *JAZ10* expression (Extended Data Fig. 8d) and wound-induced *JAZ10* expression was not reduced relative to wild type in *rbohD* mutants (Extended Data Fig. 8e) in which WASP production was similar to that in the wild type (Extended Data Fig. 9). In conclusion, the propagating signal leading to distal *JAZ10* expression is likely to be RBOHD-independent.

GLR genes mediate long-distance wound signalling

There is growing evidence that genes encoding various ion channels and pumps can affect jasmonate signalling[35,36]. For example, the overexpression of a *GLUTAMATE RECEPTOR-LIKE* (*GLR*) gene from radish stimulated the expression of jasmonate-regulated genes, including

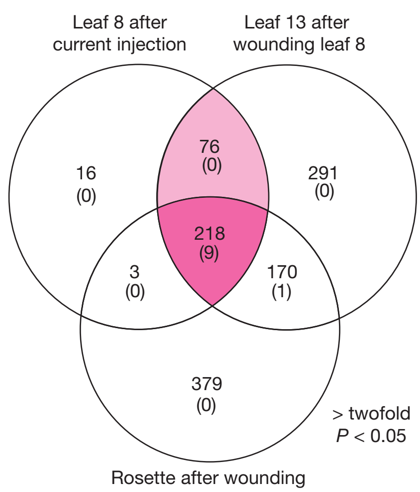


Figure 3 | Current injection and wounding stimulate the expression of a common *JAZ* gene-enriched subset of genes. Venn diagram for the number of transcripts upregulated more than twofold compared to unstimulated plants ($P \leq 0.05$) for current-injected leaf 8 (40 μA, 10 s; this study), for leaf 13 of plants that had been wounded on leaf 8 (this study), and for wounded *Arabidopsis* rosettes[32]. Numbers in parentheses are upregulated *JAZ* genes.

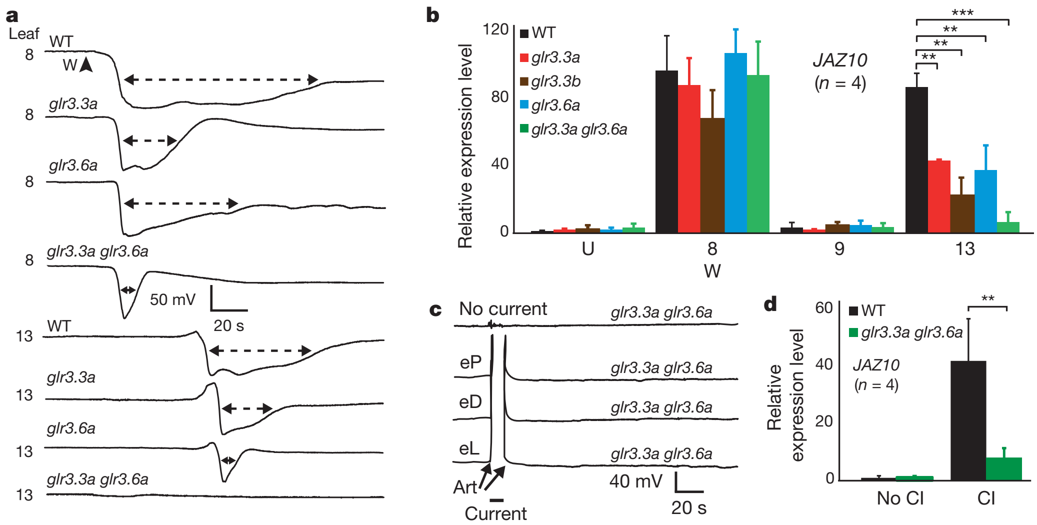


Figure 4 | *glr* mutants reduce the duration of WASPs, and responses to current injection. **a**, Typical WASPs in wild type (WT), *glr3.3a*, *glr3.6a* and *glr3.3a glr3.6a* after wounding leaf 8. Dashed line, WASP duration. **b**, *JAZ10* expression (± s.d.) after wounding leaf 8 in the WT, *glr* single mutants, and the *glr3.3a glr3.6a* double mutant. **c**, Lack of surface potential generation in leaf 8 of the *glr3.3a glr3.6a* double mutant after CI. **d**, *JAZ10* expression (± s.d.) after CI. For **b** and **d**, RNA samples were collected 1 h after stimulation. $^{**}P < 0.01$, $^{***}P < 0.001$.

VSP1, in *Arabidopsis*[37]. Using a reverse genetic approach based on monitoring electrical activity after wounding leaf 8, we screened 34 homozygous *Arabidopsis* mutants in putative pumps and ion channels including *GLRs* (Extended Data Fig. 9). Mutations in 4 genes, *GLR3.1*, *GLR3.2* (two alleles), *GLR3.3* (two alleles) and *GLR3.6* (two alleles), all of which reduced *GLR* expression (Extended Data Fig. 10), caused reduced durations of surface potential changes either in leaf 8, leaf 13, or both (Fig. 4a and Extended Data Fig. 9). Two of the *glr* mutations, one that reduced WASP duration in leaf 8 (*glr3.3a*), and one that reduced WASP duration in leaf 13 (*glr3.6a*), were combined to produce *glr3.3a glr3.6a*. This genotype, and a second related double mutant showed reduced electrical activity in wounded leaf 8 but, unlike the single mutants, changes in surface potential were no longer detectable in leaf 13 when leaf 8 was wounded (Fig. 4a and Extended Data Fig. 9). Additionally, although wounding caused elevated *JAZ10* transcript levels in wounded leaf 8 of the wild type, in the *glr* single mutants, and in the *glr3.3 glr3.6* double mutant, *JAZ10* expression was reduced in distal leaf 13 of both *glr* single mutants (Fig. 4b). A stronger reduction in wound-induced *JAZ10* transcript level relative to the wild type was seen in distal leaf 13 of the *glr3.3a glr3.6a* double mutant (Fig. 4b). Finally, we were unable to stimulate electrical activity after current injection in the double mutant (Fig. 4c) and the elevation of *JAZ10* transcript levels seen in leaf 8 of the wild type after CI was almost abolished (Fig. 4d).

Discussion

We have identified genes involved in the propagation of electrical activity leading to defence gene expression. The findings are consistent with a previous report implicating electrical signalling in the distal activation of proteinase inhibitor gene expression in tomato seedlings[5]. The *GLR* genes we studied encode putative cation channels, and *GLR3.3* functions in agonist-stimulated plasma membrane depolarization[38,39]. This gene[38], and several *GLRs* expressed in pollen[40], can control cytosolic Ca^{2+} influxes, and GLRs have also been implicated in mediating calcium influxes in response to the perception of microbe-associated molecular patterns[41]. Our results now show that GLRs control the distal wound-stimulated expression of several key jasmonate-inducible regulators of jasmonate signalling (*JAZ* genes) in the adult-phase plant. Finally, GLRs are related to ionotropic glutamate receptors (iGluRs) that are important for fast excitatory synaptic transmission in the vertebrate nervous system[42]. They and their plant relatives may control signalling mechanisms that existed before the divergence of animals and plants[43]. If so, a deeply conserved function for these genes might be to link damage perception to distal protective responses.

METHODS SUMMARY

Arabidopsis thaliana accession Col-0 and T-DNA insertion lines were obtained from the Nottingham *Arabidopsis* Stock Centre (NASC). Their homozygosity was confirmed before all experiments. Surface potentials were recorded with silver/silver chloride electrodes placed in 10 μl of 10 mM KCl in 0.5% (w/v) agar. The ground electrode was placed in the soil. Current injection was carried out via two platinum wires that were inserted into the leaf one day before the experiment. Quantitative PCR for *JAZ10* (At5g13220) and *VSP2* (At5g24770) was from reverse-transcribed RNA by SYBR Green assays and standardized to ubiquitin-conjugating enzyme (*UBC21*, At5g25760). For transcriptome analysis, amplified RNA was hybridized to Affymetrix ATH1 arrays. Probe sets showing at least a twofold change and a *P* value of < 0.05 were considered significant. Jasmonates, extracted according to ref. 10, were separated by high-performance liquid chromatography (Phenomenex Kinetex; 2.6 mm C18 100 Å column) and quantified by electrospray ionization mass spectrometry in the multiple reaction monitoring mode.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 6 August 2012; accepted 22 July 2013.

1. Card, G. & Dickinson, M. H. Visually mediated motor planning in the escape response of *Drosophila*. *Curr. Biol.* **18,** 1300–1307 (2008).
2. Escalante-Pérez, M. *et al.* A special pair of phytohormones controls excitability, slow closure, and external stomach formation in the Venus flytrap. *Proc. Natl Acad. Sci. USA* **108,** 15492–15497 (2011).
3. Walters, D. R. *Plant Defense* (Blackwell, 2011).
4. Koo, A. J. K. & Howe, G. A. The wound hormone jasmonate. *Phytochemistry* **70,** 1571–1580 (2009).
5. Wildon, D. C. *et al.* Electrical signaling and systemic proteinase-inhibitor induction in the wounded plant. *Nature* **360,** 62–65 (1992).
6. Browse, J. Jasmonate passes muster: a receptor and targets for the defense hormone. *Annu. Rev. Plant Biol.* **60,** 183–205 (2009).
7. Howe, G. A. & Jander, G. Plant immunity to insect herbivores. *Annu. Rev. Plant Biol.* **59,** 41–66 (2008).
8. Fonseca, S. *et al.* (+)-7-*iso*-Jasmonoyl-L-isoleucine is the endogenous bioactive jasmonate. *Nature Chem. Biol.* **5,** 344–350 (2009).
9. Glauser, G. *et al.* Spatial and temporal dynamics of jasmonate synthesis and accumulation in *Arabidopsis* in response to wounding. *J. Biol. Chem.* **283,** 16400–16407 (2008).
10. Glauser, G. *et al.* Velocity estimates for signal propagation leading to systemic jasmonic acid accumulation in wounded *Arabidopsis*. *J. Biol. Chem.* **284,** 34506–34513 (2009).
11. Koo, A. J. K., Gao, X., Jones, A. D. & Howe, G. A. A rapid wound signal activates the systemic synthesis of bioactive jasmonates in *Arabidopsis*. *Plant J.* **59,** 974–986 (2009).

12. Reymond, P. *et al.* A conserved transcript pattern in response to a specialist and a generalist herbivore. *Plant Cell* **16,** 3132–3147 (2004).
13. Maffei, M., Bossi, S., Spiteller, D., Mithöfer, A. & Boland, W. Effects of feeding *Spodoptera littoralis* on Lima bean leaves. I. Membrane potentials, intracellular calcium variations, oral secretions, and regurgitate components. *Plant Physiol.* **134,** 1752–1762 (2004).
14. Fromm, J. & Lautner, S. Electrical signals and their physiological significance in plants. *Plant Cell Environ.* **30,** 249–257 (2007).
15. Stahlberg, R., Cleland, R. E. & Volkenburgh, E. V. in *Communication in plants* (eds Baluška, F., Mancuso, S. & Volkmann, D.) 291–308 (Springer-Verlag, 2006).
16. Boller, T. & Felix, G. A Renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.* **60,** 379–406 (2009).
17. Krol, E. *et al.* Perception of the *Arabidopsis* danger signal peptide 1 involves the pattern recognition receptor AtPEPR1 and its close homologue AtPEPR2. *J. Biol. Chem.* **285,** 13471–13479 (2010).
18. Schaller, A. & Oecking, C. Modulation of plasma membrane H^+-ATPase activity differentially activates wound and pathogen defense responses in tomato plants. *Plant Cell* **11,** 263–272 (1999).
19. Schaller, A. & Frasson, D. Induction of wound response gene expression in tomato leaves by ionophores. *Planta* **212,** 431–435 (2001).
20. Fisahn, J., Herde, O., Willmitzer, L. & Peña-Cortés, H. Analysis of the transient increase in cytosolic Ca^{2+} during the action potential of higher plants with high temporal resolution: requirement of Ca^{2+} transients for induction of jasmonic acid biosynthesis and *PINII* gene expression. *Plant Cell Physiol.* **45,** 456–459 (2004).
21. Stahlberg, R. & Cosgrove, D. J. Comparison of electric and growth-responses to excision in cucumber and pea-seedlings. 1. Short-distance effects are a result of wounding. *Plant Cell Environ.* **17,** 1143–1151 (1994).
22. Carpaneto, A. *et al.* Cold transiently activates calcium-permeable channels in *Arabidopsis* mesophyll cells. *Plant Physiol.* **143,** 487–494 (2007).
23. Minorsky, P. V. Temperature sensing by plants: a review and hypothesis. *Plant Cell Environ.* **12,** 119–135 (1989).
24. Dengler, N. G. The shoot apical meristem and development of vascular architecture. *Can. J. Bot.* **84,** 1660–1671 (2006).
25. Yan, Y. *et al.* A downstream mediator in the growth repression limb of the jasmonate pathway. *Plant Cell* **19,** 2470–2483 (2007).
26. Chauvin, A., Caldelari, D., Wolfender, J.-L. & Farmer, E. E. Four 13-lipoxygenases contribute to rapid jasmonate synthesis in wounded *Arabidopsis* leaves: a role for LOX6 in responses to long distance wound signals. *New Phytol.* **197,** 566–575 (2013).
27. Zimmermann, M. R., Maischak, H., Mithofer, A., Boland, W. & Felle, H. H. System potentials, a novel electrical long-distance apoplastic signal in plants, induced by wounding. *Plant Physiol.* **149,** 1593–1600 (2009).
28. Favre, P. & Agosti, R. D. Voltage-dependent action potentials in *Arabidopsis thaliana*. *Physiol. Plant.* **131,** 263–272 (2007).
29. Herde, O. *et al.* Localized wounding by heat initiates the accumulation of proteinase inhibitor II in abscisic acid-deficient plants by triggering jasmonic acid biosynthesis. *Plant Physiol.* **112,** 853–860 (1996).
30. Liu, Y. *et al.* *Arabidopsis* vegetative storage protein is an anti-insect acid phosphatase. *Plant Physiol.* **139,** 1545–1556 (2005).
31. Xie, D.-X., Feys, B. F., James, S., Nieto-Rostro, M. & Turner, J. G. *COI1*: An *Arabidopsis* gene required for jasmonate-regulated defense and fertility. *Science* **280,** 1091–1094 (1998).
32. Kilian, J. *et al.* The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J.* **50,** 347–363 (2007).
33. Walley, J. W. *et al.* Mechanical stress induces biotic and abiotic stress responses via a novel *cis*-element. *PLoS Genet.* **3,** e172 (2007).
34. Miller, G. *et al.* The plant NADPH oxidase RBOHD mediates rapid systemic signaling in response to diverse stimuli. *Sci. Signal.* **2,** ra45 (2009).
35. Brüx, A. *et al.* Reduced V-ATPase activity in the trans-Golgi network causes oxylipin-dependent hypocotyl growth inhibition in *Arabidopsis*. *Plant Cell* **20,** 1088–1100 (2008).
36. Bonaventure, G. *et al.* A gain-of-function allele of TPC1 activates oxylipin biogenesis after leaf wounding in *Arabidopsis*. *Plant J.* **49,** 889–898 (2007).
37. Kang, S. *et al.* Overexpression in *Arabidopsis* of a plasma membrane-targeting glutamate receptor from small radish increases glutamate-mediated Ca^{2+} influx and delays fungal infection. *Mol. Cells* **21,** 418–427 (2006).
38. Qi, Z., Stephens, N. R. & Spalding, E. P. Calcium entry mediated by GLR3.3, an *Arabidopsis* glutamate receptor with a broad agonist profile. *Plant Physiol.* **142,** 963–971 (2006).
39. Stephens, N. R., Qi, Z. & Spalding, E. P. Glutamate receptor subtypes evidenced by differences in desensitization and dependence on the *GLR3.3* and *GLR3.4* genes. *Plant Physiol.* **146,** 529–538 (2008).
40. Michard, E. *et al.* Glutamate receptor–like genes form Ca^{2+} channels in pollen tubes and are regulated by pistil D-serine. *Science* **332,** 434–437 (2011).
41. Kwaaitaal, M., Huisman, R., Maintz, J., Reinstädler, A. & Panstruga, R. Ionotropic glutamate receptor (iGluR)-like channels mediate MAMP-induced calcium influx in *Arabidopsis thaliana*. *Biochem. J.* **440,** 355–365 (2011).
42. Traynelis, S. F. *et al.* Glutamate receptor ion channels: structure, regulation, and function. *Pharmacol. Rev.* **62,** 405–496 (2010).
43. Chiu, J. C. *et al.* Phylogenetic and expression analysis of the glutamate-receptor–like gene family in *Arabidopsis thaliana*. *Mol. Biol. Evol.* **19,** 1066–1082 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements Supported by a Faculty of Biology and Medicine Interdisciplinary grant (to S.K. and E.E.F.) and Swiss NSF grants 3100A0-122441 and 31003A-138235 (to E.E.F.). We thank I. Acosta, D. Gasperini, S. Stolz and A. Chételat and other Farmer lab members for critical comments and/or technical help, M. Blanchard for help with electrophysiology, and the Lausanne Genomic Technologies Facility and M. Shakhsi-Niaei for help with transcriptome analyses. We thank Y. Lee and F. Mauch for *rbohD* seeds, P. Schweizer and P. Reymond for insect larvae, J.-L. Wolfender for analytics support, and R. Benton, C. Fankhauser, N. Geldner, C. Hardtke and Y. Poirier for valuable comments.

Author Contributions S.A.R.M., A.C., F.P. and S.K. performed experiments; E.E.F., S.A.R.M. and S.K. conceived experiments; E.E.F. and S.K. wrote the manuscript.

Author Information Gene expression data are available in the GEO database under accession number GSE41779. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.E.F. (edward.farmer@unil.ch).

METHODS

Plant material and growth conditions and bioassays. *Arabidopsis thaliana* (Columbia) were soil-grown (one seed per 7-cm diameter pot) for 5 to 6 weeks with 10 h light (100 $\mu E\, s^{-1}\, m^{-2}$), 70% humidity; day 22 °C, night 18 °C. One single wounding experiment was carried out per plant. Wounds were inflicted with plastic non-locking thumb forceps that had flat, 4-mm-wide ridged tips. The space between each ridge was 1 mm with an inter-ridge depth of 0.6 mm. Wounds were inflicted with these ridges parallel to the long axis of the leaf. The first wound was made at the leaf tip and the second wound was made so that it abutted the first, and so on until 40–50% of the leaf was wounded. The wounding procedure takes less than 10 s to complete. Prior to electrophysiology experiments plants were moved into a Faraday cage under the same light conditions. *Spodoptera littoralis* (4th instar larvae) were placed on plants or, alternatively, the apical parts of the leaves were crushed with plastic forceps. A transparent plastic support was used to stabilize the wounded leaf during the experiments. Leaves (excluding cotyledons) were numbered from old to young.

Surface potential recordings and current injection. For surface potential recordings, silver electrodes 0.5 mm in diameter (World Precision Instruments) were chloridized with HCl (0.1 M), stored at room temperature and rechloridized after several uses. Experiments were conducted in a controlled environment room without changing the growth conditions. Two 2-channel amplifiers (FD 223 and Duo 773, World Precision Instruments) were simultaneously used to record the surface potential at four positions. The electrode–leaf interface was a drop (10 μl) of 10 mM KCl in 0.5% (w/v) agar placed so that the silver electrode did not contact and damage the cuticle. The inter-electrode distance was the distance between the nearest edges of these agar droplets. The ground electrode was placed in the soil. The procedure for data quantitation is shown in Fig. 1b in the main text. Latency is the period between wounding and WASP detection. Amplitude was relative to the baseline before wounding. Duration is the time between amplitude change midpoints. For experiments on interrupting signals, ceramic scissors (CS-250 Kyocera) were used. For current injection two platinum wire electrodes (Advent Research Materials), 0.1-mm diameter were inserted in the midrib 1 cm apart (Fig. 2a) so that the end of the wire was visible from the abaxial leaf side but did not make contact with the soil. After insertion of the Pt wires the plants were rested for 24 h before experiments. The two Pt wires were connected to a homemade current source that was controlled by the acquisition program. Current was injected between the two Pt wires and the wire closest to the leaf lamina served as the positive electrode. To optimise current injection to generate surface potential changes we used the experimental setup shown in Fig. 2a in the main text. Current was injected into the petiole and surface potentials were measured with an electrode placed on the midrib in position eD. Combinations of 10, 20 or 40 μA for 1 or 10 s were tested. Only injecting 40 μA for 10 s led to reproducible surface potential changes in the lamina. Control leaves (no CI) carried Pt wire implants but were not subjected to CI. Control measurements indicated that during the injection of 40 μA the voltage difference between the two platinum electrodes was 12.7 ± 0.9 V ($n = 9$). Surface potentials were recorded as described above. In the combined experiments, the Chartmaster program via the InstruTECH LIH 8+8 interface (HEKA Electronic) was used to record the induced surface potential changes and to control the time and duration of current injection. In the experiments without current injection, surface potentials were recorded with Datatrax2 software via the LabTrax-4/16 interface (World Precisions Instruments). The sampling interval was 10 ms. Control plants were implanted with Pt wires in all current injection experiments. Trypan blue staining[44] was used to assess the extent of damage caused by implanting Pt wires in petioles. The wires were implanted 24 h before staining. For β-glucuronidase (GUS) reporter plants the current injected leaf and the cognate control leaf were harvested for GUS staining for 15 h at 37 °C. The tissue was destained in 70% ethanol[45].

Ion leakage. To investigate the effects of current injection on ion leakage, Pt current injection wires were inserted into petioles and the plants were left overnight before current injection (40 μA, 10 s). Three surface electrodes (eP, eD and eL, placed in 0.5% w/v agar droplets (10 μl) containing 10 mM KCl) were placed on the lamina (see Fig. 2a) and used to monitor current-induced depolarizations. As controls, another set of plants were implanted with Pt wires and surface electrodes but not subjected to current injection. After current injection plants were incubated in the light for 1 h then the current-injected leaves and control leaves were cut off at the base of petiole. Before conductivity measurements the leaf surfaces were briefly washed with water to remove the agar droplets and petioles were attached so that only the part of the lamina that is harvested for *JAZ10* measurements (see Fig. 2a) came into contact with deionised water (25 ml; 3 leaves per measurement). After 20 min gentle agitation the leaves were removed and the conductivity of the water was measured at 22 °C with a Hanna Instruments EC215 conductivity metre (Distrelec). Positive controls were leaves infiltrated on either side of the midrib of the abaxial lamina, each time with 10 μl of 1% (v/v) Triton X-100, 1 h before ion leakage analysis. Negative controls were untreated leaves and an additional control for Triton X-100 infiltrations was water (25 ml) into which 20 μl of the 1% Triton-X-100 solution was added.

Pharmacological treatments. $LaCl_3$, catalase and diphenyleneiodonium (DPI) were from Sigma. Solutions were made in water; the DPI solution contained in addition 1% (v/v) dimethyl sulphoxide (DMSO). The abaxial surface of each side of the main vein of leaf 8 was infiltrated with the inhibitor (10 μl) in the part that was later wounded. 25–30 min after infiltration leaf 8 was wounded and WASPs were recorded on leaf 13. Control experiments were carried out in exactly the same way with infiltration solutions that contained only the corresponding solvent. A similar experimental design was used for *JAZ10* expression analyses. Compounds were infiltrated into leaf 8 (two 10 μl infiltrations, one each side of the abaxial lamina). 30 min after infiltration leaf 8 was wounded and RNA was harvested 1 h later.

Quantitative PCR. Total RNA was extracted with an RNeasy Plant Mini Kit (Qiagen) or with DNA-free RNA isolation protocols[46]. Total RNA (1 μg) was copied into complementary DNA with M-MLV Reverse Transcriptase, RNase H Minus, Point Mutant (Promega) first-strand synthesis system and oligo(dT) primers according to the manufacturer's instructions. Quantitative PCR (qPCR) analysis was performed on 100 ng of cDNA in a final volume of 20 μl according to the FullVelocity SYBR Green instruction manual (Stratagene) or with a homemade master mix containing GoTaq polymerase (Promega) and its buffer, 0.2 mM dNTPs, 2.5 mM $MgCl_2$, ROX dye and SYBR green in a final volume of 20 μl. qPCR was performed in an Mx3005P spectrofluorometric thermal cycler (Stratagene). The data were calibrated to unwounded wild type. Ubiquitin-conjugating enzyme (*UBC21*) At5g25760[47] was used as reference gene. The thermal cycle conditions were: an initial denaturation at 95 °C for 2 min, followed by 40 cycles of 20 s at 95 °C, 30 s at 60 °C and 45 s at 72 °C. Three or four biological replicates were used for each experiment. Primers used were: *UBC21* (At5g25760) forward 5′-CAGTCTGTGTGTAGAGCTATCATAGCAT-3′, reverse 5′-AGAAGATTCCCTGAGTCGCAGTT-3′; *JAZ10* (At5g13220) forward 5′-ATCCCGATTTCTCCGGTCCA-3′, reverse 5′-ACTTTCTCCTTGCGATGGGAAGA-3′; *VSP2* (At5g24770) forward 5′-CCGTGTGCAAAGAGGCTTA-3′, reverse 5′-CACAACTTCCAACGGTCAC-3′; *GLR3.1* (At2g17260) forward 5′-GGCCAAGAATTCACCAGATGC-3′, reverse 5′-GACCAAGAATCGCGGTTGACA-3′; *GLR3.2* (At4g35290) forward 5′-ATTCACCAGAAGTGGCTGGG-3′, reverse 5′-TGAAGCTGTCCGGTTTCTGA-3′; *GLR3.3* (At1g42540) forward 5′-CGACCTTTCAACCGTCTTAT-3′, reverse 5′-TCGAGAAGCTAAACCAGAGAA-3′; *GLR3.6* (At3g51480) forward 5′-GATTAGAAGTGGGTTGGGGGA-3′, reverse 5′-GAGGCAATGGTGGAGGAAGT-3′.

Transcriptomics. Total RNAs from leaves were isolated and purified with RNeasy Plant Mini Kit (Qiagen). All RNA quantities were assessed with a NanoDropND-1000 spectrophotometer and the RNA quality was assessed using RNA 6000 NanoChips with the Agilent 2100 Bioanalyzer (Agilent). For each sample, 300 ng of total RNA were amplified using the MessageAmp II-Biotin Enhanced Single Round aRNA Amplification Kit (AM1791, Ambion). 12.5 μg of the resulting biotin-labelled complementary RNA was chemically fragmented. Affymetrix ATH1 (batch 1211501) arrays (Affymetrix) were hybridized with 11 μg of fragmented target, at 45 °C for 17 h and washed and stained according to the protocol described in Affymetrix GeneChip Expression Analysis Manual (Fluidics protocol FS450_0007). The arrays were scanned using the GeneChip Scanner 3000 7G (Affymetrix) and raw data was extracted from the scanned images and analysed with the Affymetrix Power Tools software package (Affymetrix). Statistical analysis was performed using the free high-level interpreted statistical language R and various Bioconductor packages (http://www.Bioconductor.org). Hybridization quality was assessed using the Expression Console software (Affymetrix). Normalized expression signals were calculated from Affymetrix CEL files using the RMA normalization method. Differential hybridized features were identified using the Bioconductor package "limma" that implements linear models for microarray data[48]. The *P* values were adjusted for multiple testing with Benjamini and Hochberg's method to control the false discovery rate (FDR)[49]. Probe sets showing at least twofold change and a P-value < 0.05 were considered significant.

JA and JA-Ile quantification. Isopropanol and methanol were obtained from VWR Prolabo and used for extraction analysis. Liquid chromatography mass spectrometry-grade acetonitrile and water from Biosolve were used for the high-performance liquid chromatography. The internal standards used were $[^{18}O]_2$jasmonic acid[50] and $[^{13}C]_6$jasmonoyl L-isoleucine[51]. Frozen leaves (200 mg, from 5-week-old plants) were ground in a ball mill extractor with internal standards (40 ng ml^{-1}) before extraction with isopropanol[10]. Chlorophyll was removed with a C18 solid-phase extraction cartridge using methanol:H_2O (85:15, v/v) for elution. The eluate was concentrated and dissolved in 100 μl methanol:H_2O (85:15, v/v). Separation was carried out on a Phenomenex Kinetex 2.6-mm C18 100-Å column (100 × 3.0 mm). Gradient elution was at a flow rate of 0.4 ml min^{-1} with the following solvent system: A = 0.1% formic acid/water, B = 0.1% formic acid/acetonitrile; 5% B for 3 min, 5–75%

B in 11 min, 75–95% B in 2 min, 95% B for 2 min and 95–5%B in 2 min. The electrospray ionisation conditions were as follows: capillary voltage 3,300 V; cone voltage 24 V; extractor 3 V; radio frequency (RF) lens 0 V; source temperature 120 °C; desolvation temperature 350 °C; cone gas flow 900 l h^{-1} and desolvation gas flow 27 l h^{-1}. Jasmonates were monitored with quantitative multiple reaction monitoring (MRM) in a Quattro micro API mass spectrometer (Waters) with an electrospray ionization interface coupled with the Agilent LC system (Hewlett Packard). Detection was performed in negative ion mode over an *m/z* range of 100–1,000. The MRM transitions were: JA: 209.1 > 58.7, [$^{18}O_2$]JA: 213.1 > 62.8, JA-Ile: 322.2 > 130.0 and JA-[^{13}C]$_6$Ile: 328.2 > 136.0 (parent > daughter). The limit of quantification (LOQ = 3× limit of detection) could reach up to 9.2 pmol g^{-1} fresh weight (FW) for JA and 4.5 pmol g^{-1} FW for JA-Ile. Data below LOQ were considered as non-informative.

Genotyping of T-DNA insertion lines. T-DNA insertion lines were obtained from the Nottingham *Arabidopsis* Stock Centre (NASC) except for the respiratory burst oxidase homologue D mutant *rbohD* (Salk_070610) which was from Y. Lee (T-DNA line) and F. Mauch (*dSpm* line[52]). For genotyping, 5 mg fresh leaf samples were placed into 96-well microtitre plates and tissues were ground using a Qiagen TissueLyser II (Retsch Technology). Then, 60 μl of extraction buffer (200 mM Tris HCl pH 7.5, 250 mM NaCl, 25 mM Na_2EDTA, 0.5% SDS) was added to each well and the samples were centrifuged at 4,000*g* for 10 min. The supernatants were transferred into new microtitre plates and the same volume (50 μl) of isopropanol was added. The plates were then centrifuged at 4,000*g* for 5 min. The resultant pellets were washed with 70% ethanol (150 μl) and centrifuged at 4,000*g* for 5 min. Finally, DNA was resuspended in 50 μl deionized water. 2 μl of this extracted DNA was used as template for each final 20 μl PCR reaction. Sequences of primer pairs used for genotyping of T-DNA insertion lines:

glr1.1 (At3g04110, salk_057748) forward, 5′-ACCTCTTGACGCGTATGAAAG-3′; reverse, 5′-GTGAAAAAGAAAAGCCAAGGG-3′. *glr1.4* (At3g07520, salk_129955) forward, 5′-TATATTTGGCCAAGCTCAACG-3′; reverse, 5′-CTTATAGTGCGGGCTTTGTTG-3′.

glr2.3 (At2g24710, salk_113260) forward, 5′-TATTTGCGGAAGTTCCATTTG-3′; reverse, 5′-AGAGCGACAAGAAACAGAACC-3′. *glr2.7* (At2g29120, salk_121990) forward, 5′-GGAAATCTTGCCGGTTAAAAG-3′; reverse, 5′-ACAAATTTGGGGACATTAGGG-3′.

glr2.8 (At2g29110, salk_111695) forward, 5′-GAGTACCTTTCCCTGACCCTG-3′; reverse, 5′-GAAGGGAGGAGAAGAATGGTG-3′. *glr2.9* (At2g29100, salk_125496) forward, 5′-TGACAAGGTGCTCCCATTATC-3′; reverse, 5′-AGAAATTCATGGTGACGGTTG-3′.

glr3.1 (At2g17260, salk_063873) forward, 5′-AGATGAACAAACGTGACCACC-3′; reverse, 5′-TGGCTTTTTGTGGTTCTGATC-3′. *glr3.2a* (At4g35290, salk_150710) forward, 5′-TTTTGGATCCAGCATTAGTCG-3′; reverse, 5′-TTTTGCGGTTTTGTTTGTAGG-3′.

glr3.2b (At4g35290, salk_133700) forward, 5′-TCCATTACTCAATTTCGGTGG-3′; reverse, 5′-AAACCCAAACCAAAATCATCC-3′. *glr3.3a* (At1g42540, salk-099757) forward, 5′-GATGCTGCATATGGTTGTGTG-3′; reverse, 5′-GTTGAACGATAAGCTTGCGAG-3′.

glr3.3b (At1g42540, salk_077608) forward, 5′-TGCTGTTGATCTCTTGCAATG-3′; reverse, 5′-CACACAACCATATGCAGCATC-3′. *glr3.4* (At1g05200, salk_079842) forward, 5′-GGGTTAATCCGGCTTATGAAG-3′; reverse, 5′-GAAGTGAGACTGGCCGTGTAG-3′.

glr3.5 (At2g32390, salk_035264) forward, 5′-TGAAGTTGCTGCAAATGTGAG-3′; reverse, 5′-TGTCGACATGTCCACAGCTAG-3′. *glr3.6a* (At3g51480, salk_091801) forward, 5′-TTCGTTCAAAGGTGGCATAAC-3′; reverse, 5′-CGACTATGAGGAAAGACGCAG-3′.

glr3.6b (At3g51480, salk_035353) forward, 5′-ATAGTCGGTGCTGTCATTTGG-3′; reverse, 5′-TCCCCAAAAGCTCTTAAGCTC-3′. *cngc12* (At2g46450, salk_092622) forward, 5′-ATTGATGCATTGAAGTCAGGG-3′; reverse, 5′-TACTTTGGTTTCGAAGCTTGC-3′.

cngc18 (At5g14870, sail_191_H04) forward, 5′-GTTTATCGCCAAGACTGCTTG-3′; reverse, 5′-TAGCATCTCATTCACCGGATC-3′. *cngc20* (At3g17700, salk_129133) forward, 5′-AAAACAGTTACCTGGAAGCCC-3′; reverse, 5′-TGCCTTTACACCACCTTTTTG-3′.

aca11 (At3g57330, salk_121482) forward, 5′-TTGCCTCACAAATTACGTTTTG-3′; reverse, 5′-ACAAACTCCCACGTTTGACAG-3′. *clc-b* (At3g27170, salk_027349) forward, 5′-TCAACCCGTGGAGTTCTGTAG-3′; reverse, 5′-GGAATTCTTGGGAGCCTGTAC-3′.

clc-e (At4g35440, salk_142812) forward, 5′-ACAAAGAACAAAAATTGGCCC-3′; reverse, 5′-CTCAACCAATCTGAGGAGCTG-3′. *kab1* (At1g04690, salk_030039) forward, 5′-GAGGGAATAGCTCCCTTGTTG-3′; reverse, 5′-GATGTGAAAGAAGCGAAATCG-3′.

akt6 (At2g25600, salk_136050) forward, 5′-GAGAGGAAGAAGAAGCCTTGC-3′; reverse, 5′-ATGGTCAGCAACATCATCCTC-3′. *skor* (At3g02850, salk_097435) forward, 5′-CCCATATCTCACTGGTTCACC-3′; reverse, 5′-CCAAACTTCAGCGAAACAGAG-3′.

tpk1 (At5g55630, salk_146903) forward, 5′-AAATGTCGAGTGATGCAGCTC-3′; reverse, 5′-TCAAGTTGCTCGAACTCATCC-3′. *tpk3* (At4g18160, salk_049137) forward, 5′-ATTGATTACAGCCATTGCTGG-3′; reverse, 5′-CCGTATATCTCCATTCGGAAC-3′.

annat6 (At5g10220, salk_043207) forward, 5′-TTCTATCCACTGTAGACAGCCTG-3′; reverse, 5′-AATACGCATCTCTCTCCGTTG-3′. *pen3* (At1g59870, salk_110927) forward, 5′-GCGAGAGTTGGACTCACTTTG-3′; reverse, 5′-TCACCCAACTAAATCCTCACG-3′.

vha-e1 (At4g11150, salk_019365) forward, 5′-AAGAGTTGGTCCTTGGAAAGC-3′; reverse, 5′-GTAGATCGGATTTTCACGACG-3′. *vha-g* (At3g01390, salk_087613) forward, 5′-GCTGTTACAATCGCTGAAAGC-3′; reverse, 5′-TTGAGCTTCTACCTCAGCAGC-3′.

vha-a2 (At2g21410, salk_142642) forward, 5′-ACCTCTGGCTCAAAATTGTCC-3′; reverse, 5′-TCCACATGAATATAGCCCGAG-3′. *aha1* (At2g18960, salk_118350) forward, 5′-TTCGATTCTCCCACACAGATC-3′; reverse, 5′-ACGGATTGTGATTGAGACTGC-3′.

aha2 (At4g30190, salk_073730) forward, 5′-GCGAAAACATATGAACTTTCGAC-3′; reverse, 5′-CTTAGGGAGCTGCACACACTC-3′. *aha3* (At5g57350, sail_810_C08) forward, 5′-GTAGATTGCAACGGCTATTGC-3′; reverse, 5′-TTGTCGTGAAGAAGCTATGGC-3′.

aha11 (At5g62670, salk_152723) forward, 5′-ATGACAGCGATTGAGGAAATG-3′; reverse, 5′-GGCAAAACAACATCATTGATG-3′. *rbohD* (At5g47910, salk_070610) forward, 5′-TTTCAACGCCTTTTGGTACAC-3′; reverse, 5′-GTTACCTATTCTTTTGCCGGG-3′.

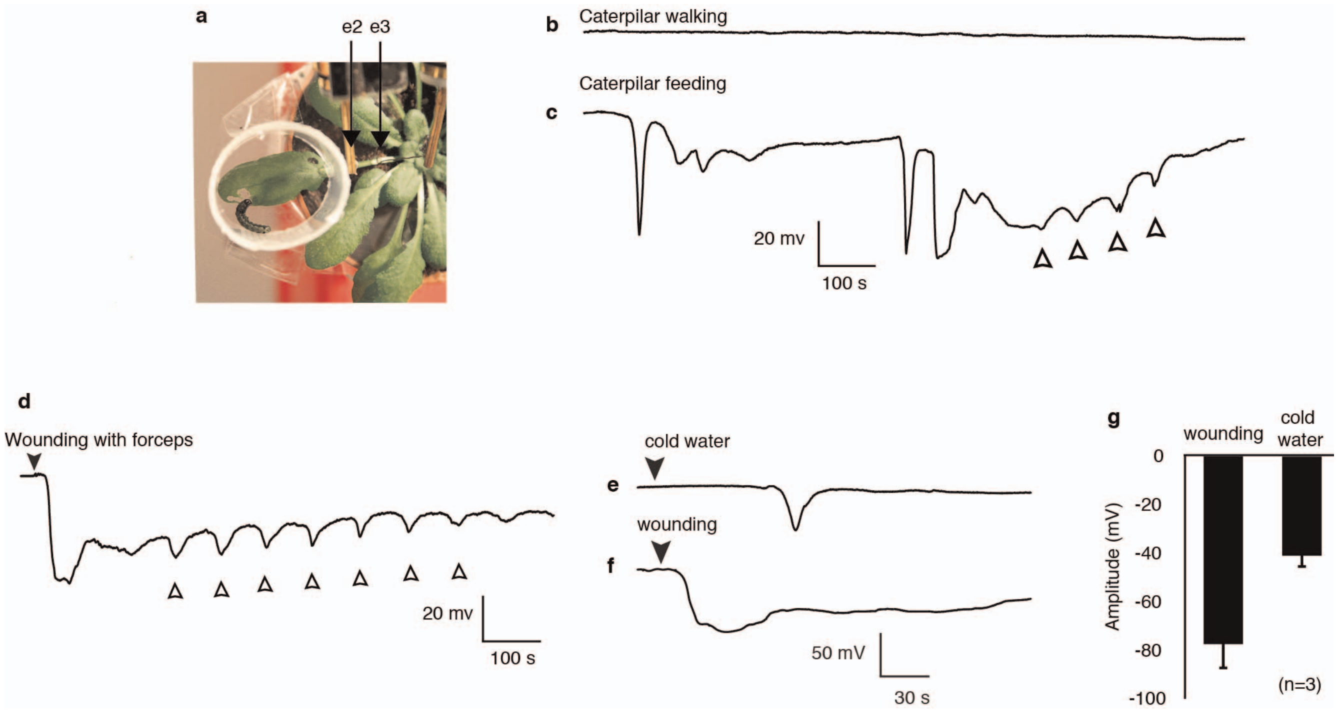
RT–PCR analysis of *glr3.3* and *glr3.6* mutants. Total RNA was extracted with DNA-free RNA isolation protocols[46]. Total RNA (1 μg) was copied into cDNA with M-MLV Reverse Transcriptase, RNase H Minus, Point Mutant first strand synthesis system (Promega, Madison WI) and oligo(dT) primers according to the manufacturer's instructions. Ubiquitin-conjugating enzyme (UBC21) At5g25760[47] was used as the reference gene. Three biological replicates were used for each experiment. Primers used were: for *glr3.3a*, forward 5′-GTTGAACGATAAGCTTGCGAG-3′ and reverse 5′-GATGCTGCATATGGTTGTGTG-3′ and for *glr3.3b* forward 5′-CACACAACCATATGCAGCATC-3′ and reverse 5′-TGCTGTTGATCTCTTGCAATG-3′. For *glr3.6a* forward 5′-TTCGTTCAAAGGTGGCATAAC-3′ and reverse 5′-AGTTGCAGCGACTTGAACCA-3′.

***VSP2*$_{Pro}$*:GUSPlus* plant transformation.** The *VSP2* (At5g24770) promoter, amplified using 5′-TTCTCTCTGGTTATATTTTGTTGCTG-3′ and 5′-TGTTTATATGTGTGACGCAAAGG-3′ primers) was cloned with XmaI and KpnI (New England Biolabs) into the pUC57-L4-KpnI/XmaI-R1 plasmid producing a pEN-L4-*VSP2*$_{Pro}$-R1 as an pENTRY clone. The pUC57-L4-KpnI/XmaI-R1 plasmid was generated by J. Vermeer by introducing L4-KpnI/XmaI-R1 *att* recombination and restriction sites into pUC57 (Invitrogen). pEN-L1-*GUSPlus*-L2 plasmids were obtained with Gateway technology according to manufacturer instructions (Invitrogen) with *GUSPlus* cDNA (amplified from pCAMBIA1305.2 (CAMBIA) and pDONR/ZEO (Invitrogen). The final *VSP2*$_{Pro}$-*GUSPlus* constructs were generated by using a double Gateway reaction into pEDO097pFR7m24GW. pEDO097pFR7m24GW was generated by inserting the FAST (fluorescence-accumulating seed technology) cassette[53] into pH7m24GW (Invitrogen) by E. M. N. Dohmann. Wild-type plants were transformed using *Agrobacterium tumefaciens* cells as described previously[54]. Transformed seeds expressing red fluorescence protein (RFP) were selected by florescence microscopy. The T_1 generation was used for experiments.

Statistics. All results in main figures and extended data with error bars are represented as mean ± s.d. according to standard methods using Microsoft Excel. Standard error was used for Extended Data Fig. 8d. The *P* values were generated with Student's one-tail unpaired *t*-tests except for Fig. 4d for which a two-tail unpaired was used. For qRT–PCR experiments, three technical replicates were used (two technical replicates for Extended Data Fig. 4c) and the biological replicates were indicated as '*n*' in the figures. The technical replicates that had ≥ 0.5 difference from mean C_t were excluded.

44. van Wees, S. Phenotypic analysis of *Arabidopsis* mutants: trypan blue stain for fungi, oomycetes, and dead plant cells. *Cold Spring Harb Protoc.* http://dx.doi.org/10.1101/pdb.prot4982 (2008).
45. Jefferson, R. A., Kavanagh, T. A. & Bevan, M. W. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J.* **20,** 3901–3907 (1987).
46. Oñate-Sánchez, L. & Vicente-Carbajosa, J. DNA-free RNA isolation protocols for *Arabidopsis thaliana*, including seeds and siliques. *BMC Res. Notes* **1,** 93–100 (2008).
47. Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K. & Scheible, W.-R. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.* **139,** 5–17 (2005).

48. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3,** e3 (2004).
49. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).
50. Mueller, M. J., Mène-Saffrané, L., Grun, C., Karg, K. & Farmer, E. E. Oxylipin analysis methods. *Plant J.* **45,** 472–489 (2006).
51. Kramell, R., Schneider, G. & Miersch, O. Chiral separation of amide conjugates of jasmonic acid by liquid chromatography. *Chromatographia* **45,** 104–108 (1997).
52. Torres, M. A., Dangl, J. L. & Jones, J. D. *Arabidopsis* gp91^{phox} homologues *AtrbohD* and *AtrbohF* are required for accumulation of reactive oxygen intermediates in the plant defense response. *Proc. Natl Acad. Sci. USA* **99,** 517–522 (2002).
53. Shimada, T. L., Shimada, T. & Hara-Nishimura, I. A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of *Arabidopsis thaliana*. *Plant J.* **61,** 519–528 (2010).
54. Berberich, T., Takahashi, Y., Saitoh, H. & Terauchi, R. in *The Handbook of Plant Functional Genomics* Ch. 6 (eds Kahl, G. & Meksem K.) 113–136 (Wiley, 2008).



Extended Data Figure 1 | Insect- and mechanical-damage-induced membrane depolarizations. **a**, The setup showing the ring cage around the insect (*S. littoralis*) and the position of the recording electrodes (e2 and e3) on leaf 8. **b**, Surface potential recording from electrode e2 while *S. littoralis* walked on the leaf. **c**, Typical surface potential changes recorded on electrode e3 during *S. littoralis* feeding. The arrowheads indicate periodicity in the signal. **d**, A proportion of WASPs induced by mechanical damage show periodicity. Filled arrowhead, time of wounding. The apical 40% of leaf 8 was wounded with forceps. Periodicity (unfilled arrowheads) was seen in 61% ($n = 110$) of experiments. **e**, Chilling-induced depolarization generated by gently placing water (150 μl, 0 °C) onto leaf 8 at the time indicated with the arrowhead. Chilling induced a change in surface potential in 3 out of 7 recordings. **f**, Typical WASP of the same polarity. For **d**, **e** and **f** the recording electrode was on leaf 8 at position e3 (Fig. 1a in the main text). **g**, Amplitude of the change in surface potential (± s.d.) induced by wounding or by cold water.

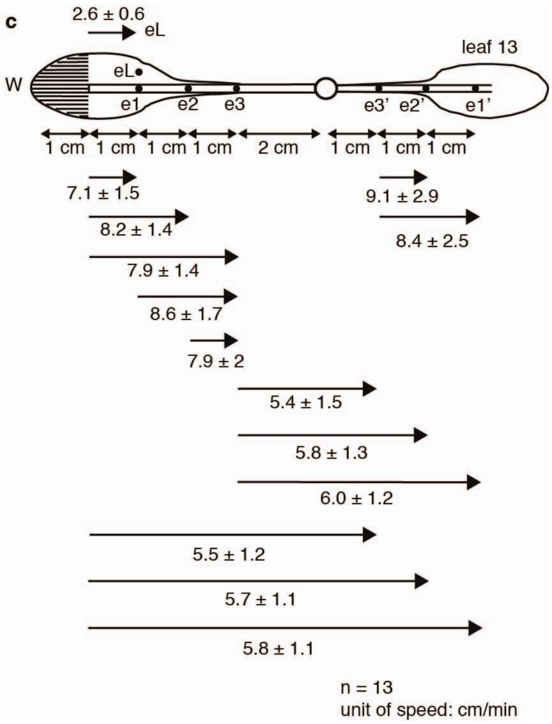
a

Electrodes	Latency (s)	Amplitude (mV)	Duration (s)	Velocity (cm/min)	x/n
Lamina (eL)	24 ± 7	-73 ± 23	96 ± 14	2.6 ± 0.6	13/15
Midrib (e1)	9 ± 3	-71 ± 15	143 ± 22	7.1 ± 1. 5	13/15
Petiole/midrib junction (e2)	15 ± 3	-73 ± 14	164 ± 32	8.2 ± 1.4	15/15
Petiole (e3)	23 ± 5	-71 ± 13	160 ± 28	7.9 ± 1.4	15/15
Average		-72 ± 16			

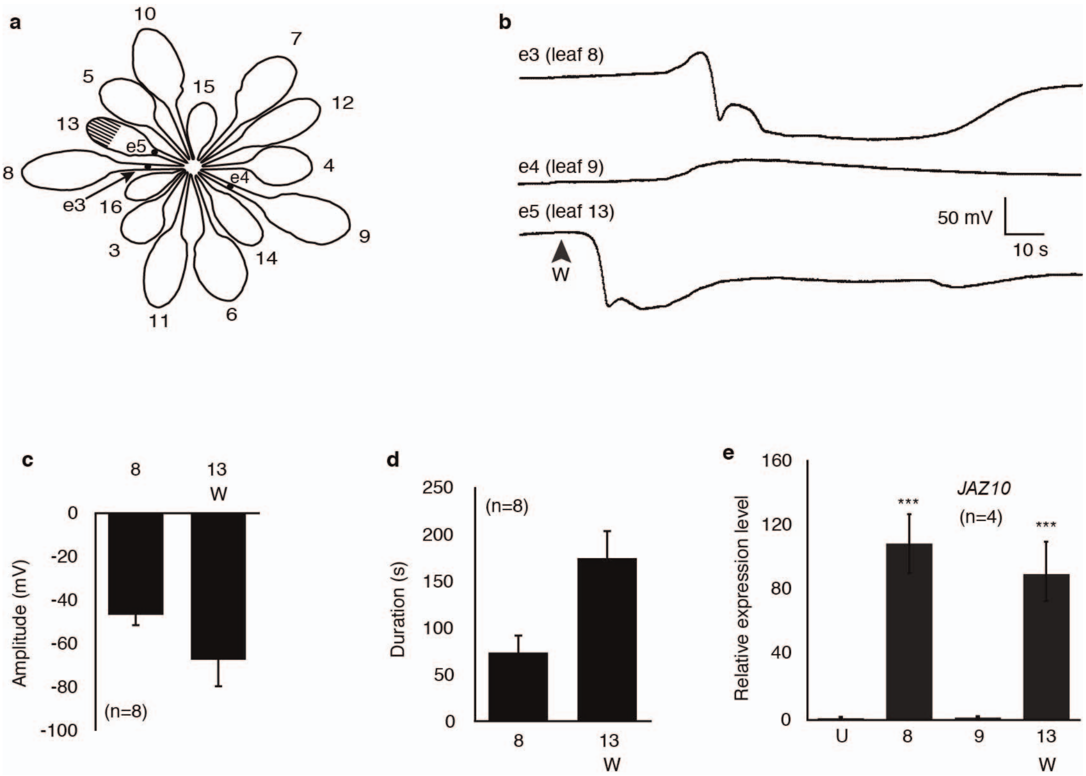
b

Electrodes	Latency (s)	Amplitude (mV)	Duration (s)	x/n
Leaf 5	66 ± 14	-57 ± 14	64 ± 16	26/27
Leaf 8 W	22 ±3	-76 ± 11	163 ± 30	33/33
Leaf 9	54 ± 12	20 ± 5	21 ± 9	46/46
Leaf 11	69 ± 14	-56 ± 13	36 ± 18	41/48
Leaf 13	66 ± 13	-51 ± 9	78 ± 20	61/61
Leaf 16	73 ± 14	-45 ± 10	56 ± 21	24/27

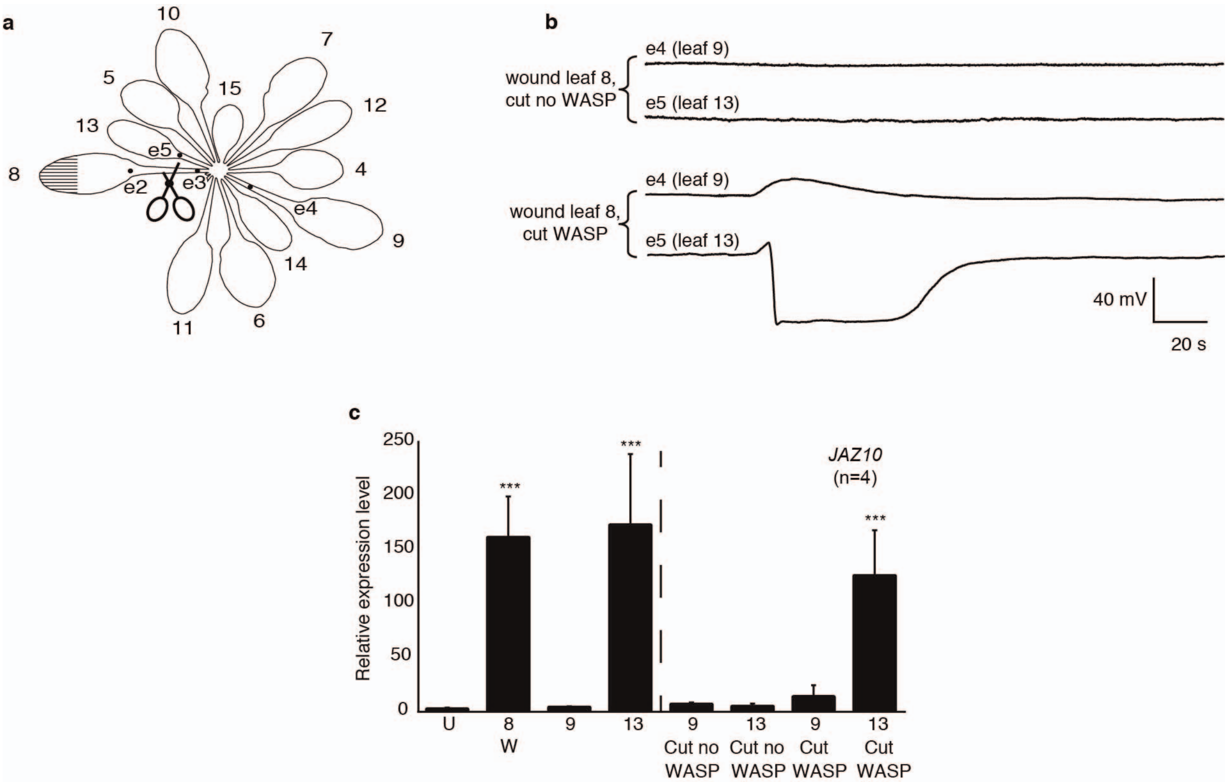
c



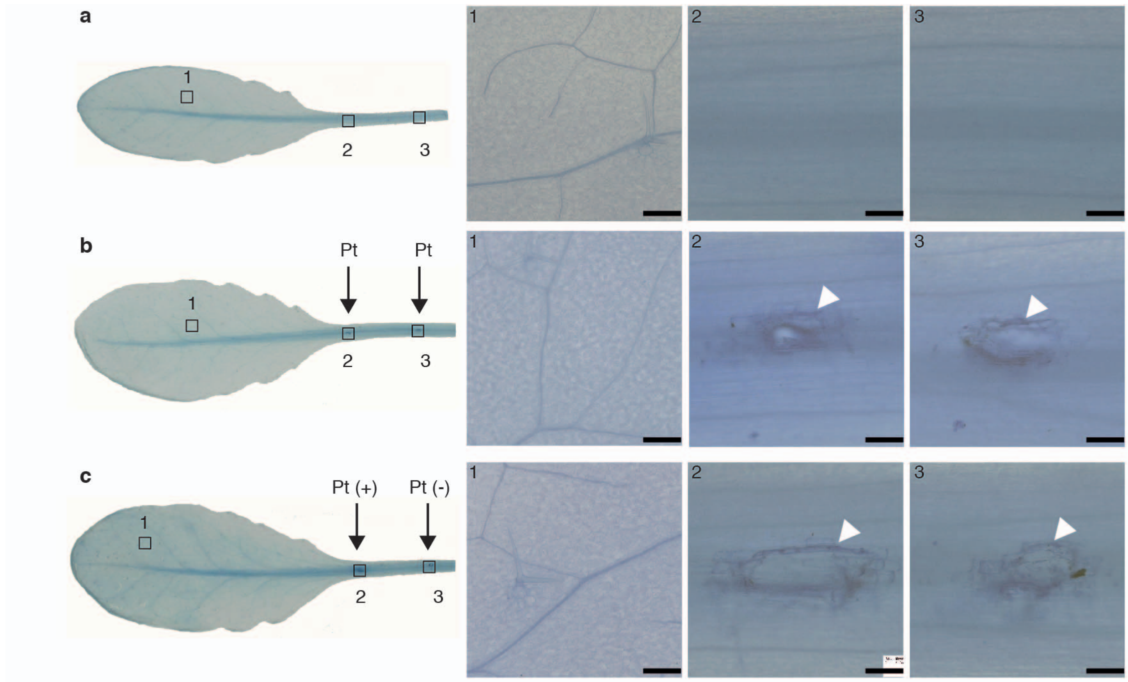
Extended Data Figure 2 | Apparent heterogeneity in WASP velocities. **a**, WASP characteristics in wounded leaf 8. **b**, Wound-activated surface potential changes in leaves 5, 9, 11, 13 and 16. Leaf 8 was wounded and surface potentials were monitored in distal leaves with electrodes placed on these leaves at position e3′. For leaf 8 the monitoring electrode was at position e2. W, wounded; *x*, number of experiments in which amplitudes of surface potentials exceeded −10 mV. Values are means ± s.d. **c**, Leaf-to-leaf signal speeds. Leaves 8 or 12 (the largest rosette leaves in 6-week-old plants) were chosen for estimating the apparent velocities of signals that travel within the wounded leaf. For leaf-to-leaf recordings, leaf 8 was wounded and recordings were made both on this leaf and on leaf 13. Analysis of variance (ANOVA) followed by Bonferroni post-hoc test showed that the WASP speed indicated in cm min^{-1} along the midrib and petiole within a leaf was not significantly different between leaves 8, 12 and 13, but was faster than the overall signalling speed from leaf 8 to leaf 13, and the signal speed from the wound to the lamina electrode (eL).

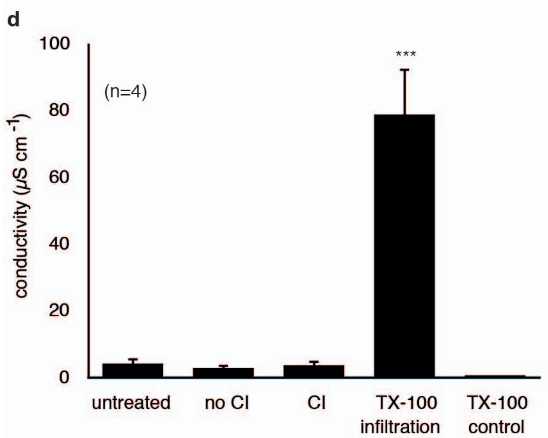


Extended Data Figure 3 | Wounding young leaves triggers WASPs and *JAZ10* expression in older leaves. **a**, Electrode placements on leaves 8 (e3), 9 (e4) and 13 (e5). **b**, Typical changes in surface potential in leaves 8, 9 and 13 after wounding leaf 13. Arrowhead shows the time of wounding (W). **c**, WASP amplitudes (± s.d.) after wounding of leaf 13. **d**, WASP durations (± s.d.) after wounding of leaf 13. **e**, *JAZ10* expression 1 h after wounding leaf 13 (± s.d.). U, unwounded leaves; W, wounded leaf 13. ***$P < 0.001$ (± s.d.).



Extended Data Figure 4 | Effects of interrupting WASP propagation on *JAZ10* expression. **a**, Experimental design: electrodes were placed on the midrib (e2) and petiole base (e3) of leaf 8, and on leaf 9 (e4) and leaf 13 (e5). 40% of leaf 8 was wounded. **b**, WASP traces for leaves 9 (non-parastichious) and leaf 13 (connected) provoked by wounding leaf 8. The first pair of traces was recorded when leaf 8 was severed upon detection of a signal at e2 and before a WASP was detected at e3. The second pair of traces was recorded when the WASP generated by wounding leaf 8 was allowed to reach e3 and the leaf was then severed immediately. **c**, *JAZ10* expression in unwounded leaves (U), wounded leaf 8 (W) and leaves 9 and 13. Left of dashed line: *JAZ10* levels in leaves 8, 9 and 13 of intact control plants 1 h after wounding leaf 8. Right of the dashed line: plants in which the wounded leaf 8 was severed when WASPs were detected at e2 but were not allowed to reach electrode e3 (cut no WASP) or when WASPs were allowed to reach e3 before severing leaf 8 (cut WASP). $^{***}P < 0.001$ ($\pm$ s.d.). Note: compared to crush-wounding, severing the petioles of otherwise undamaged leaves with sharp blades does not activate jasmonate signalling strongly in distal leaves.



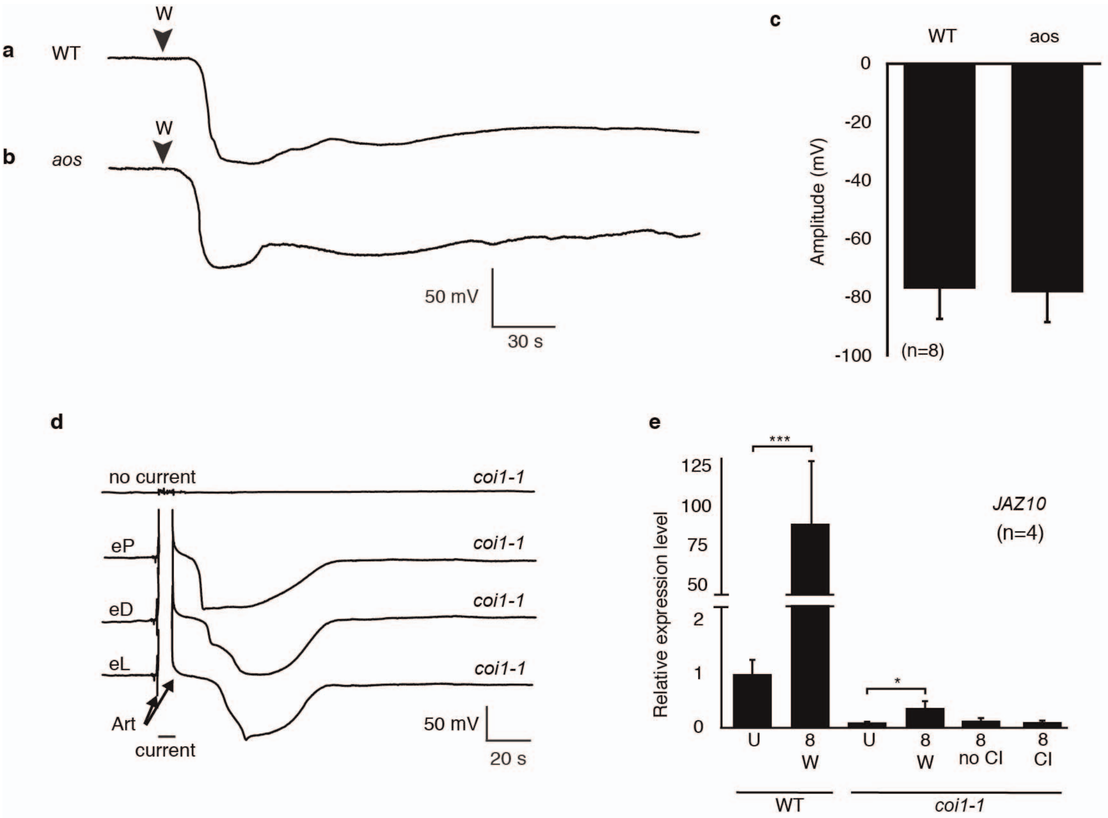


e

Electrodes	Latency (s)	Amplitude (mV)	Duration (s)	x/n
Lamina (eL)	22 ± 9	-87 ± 21	47 ± 21	33 /47
Midrib (eP)	6 ± 3	-79 ± 12	75 ± 20	44/47
midrib (eD)	15 ± 3	-80 ± 24	52 ± 23	43/47
Average		-81 ± 19	59 ± 25	

Note: An apparent velocity of surface potential displacement of 6.4 ± 1.9 cm min^{-1} was estimated from recordings at eP and eD.

Extended Data Figure 5 | Current injection does not cause cell death in the lamina but elicits surface potential changes. a–c, Trypan blue staining. **a**, Undamaged leaf. **b**, Pt wires inserted but no current injected. **c**, Current-injected leaf. Leaves were harvested 1 h after current injection. Cells were killed around the Pt wires (arrowheads) but CI did not cause increased staining of the lamina. Scale bars in boxes, 200 µm. **d**, Ion leakage analysis after current injection (CI). For controls leaves were either untreated or implanted with Pt wires and connected to three surface electrodes on the laminas (no CI). A further set of leaves was prepared identically but subjected to CI (40 µA, 10 s; CI) and harvested 1 h later for conductivity analyses. Positive controls: leaves infiltrated with 20 µl Triton X-100 (1% v/v in water) 1 h before harvest ('TX-100 infiltration'). For analysis, leaves were excised at the base of the petiole and attached so that only their laminas were bathed in deionised water (25 ml) for 20 min at 22 °C. A control for the Triton X-100 infiltration was 20 µl Triton X-100 (1% v/v in water; TX-100 control), ± s.d. **d**, Surface potential changes in different parts of leaf 8 generated by current injection. Current (40 µA, 10 s) was injected into the petiole of leaf 8 (see Fig. 2a in the main text for electrode placements). x/n = the number of experiments in which signal amplitudes exceeded −10 mV/total number of experiments. Values are means ± s.d.

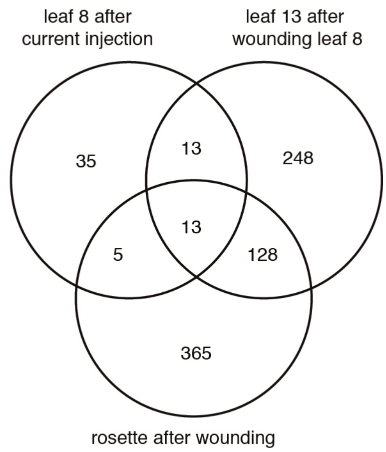


Extended Data Figure 6 | WASP generation in jasmonate biosynthesis and perception mutants. **a**, Typical recording from leaf 8 of the wild type after wounding the leaf tip. **b**, A typical recording from leaf 8 of the *allene oxide synthase* (*aos*) mutant after similar damage. In both cases the recording electrode was placed at position e3 (shown in Fig. 1a in the main text) before wounding the apical 40% of leaf 8. Arrowheads indicate the time of wound infliction (W). **c**, WASP amplitude (± s.d.) in wild-type and *aos* plants. **d**, Surface potential changes following CI (40 μA for 10 s) in the *coronatine-insensitive 1-1* (*coi1-1*) mutant. Art, artefacts recorded in the leaf during CI (bar = 10 s). Note that the signal amplitude at eP reaches a maximum before that at eD and eL. For electrode placements see Fig. 2a. **e**, Relative *JAZ10* levels in wounded WT and in the *coi1-1* mutant that had been wounded or into which current (40 μA, 10 s) had been injected. Leaves were harvested 1 h after wounding or current injection. U, unwounded; W, wounded; CI, current injection. Significant differences from the unwounded wild type are indicated, $*P < 0.05$, $***P < 0.001$ (± s.d.).

a

locus	annotation	CI / no CI		Leaf 13/control		W/control	
		FC	*P*-value	FC	*P*-value	FC	*P*-value
At1g19180	*JAZ1*	4.7	8.4E-07	61.5	3.6E-12	22.7	5.3E-04
At1g74950	*JAZ2*	5.5	5.1E-08	19	4.6E-11	8	1.6E-06
At3g17860	*JAZ3*	2.4	1.8E-06	5.9	3.9E-10	5.5	4.8E-06
At1g17380	*JAZ5*	13.6	6.4E-09	102.8	3.6E-12	22	4.7E-06
At1g72450	*JAZ6*	4.3	3.9E-08	12.8	2.7E-11	5	5.7E-06
At2g34600	*JAZ7*	25.9	9.0E-12	143	3.2E-14	18.6	7.5E-04
At1g30135	*JAZ8*	6.9	1.6E-07	51	1.9E-11	16.8	5.8E-06
At1g70700	*JAZ9*	7	1.1E-07	24.8	2.0E-10	6.1	1.6E-05
At5g13220	*JAZ10*	28.2	1.9E-10	123.1	1.5E-12	28.8	8.4E-06
At5g20900	*JAZ12*	1.8	1.9E-05	2.9	2.0E-08	2.5	6.2E-05

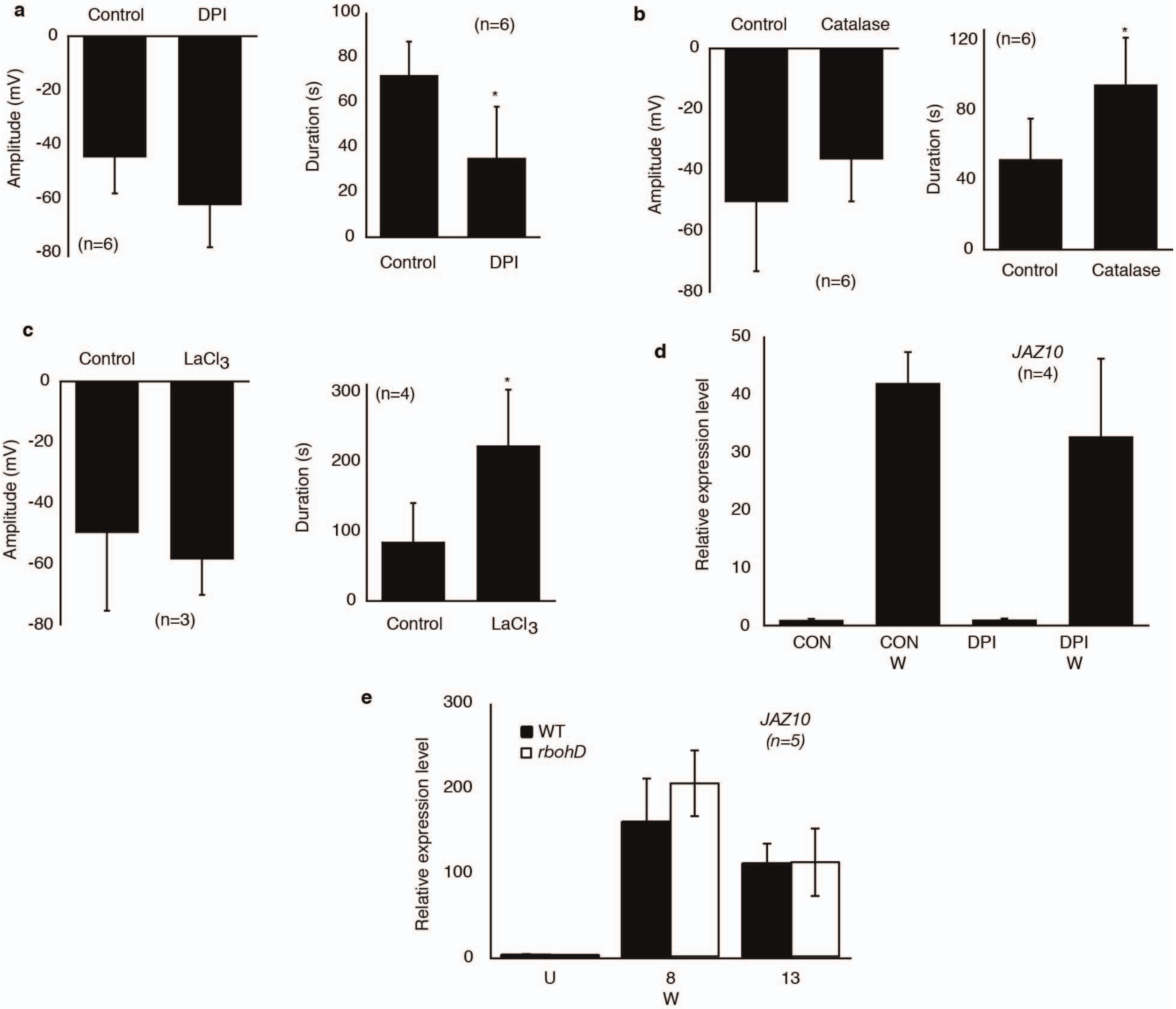
b



c

locus	annotation	CI / no CI		Leaf 13/control		W/control	
		FC	*P*-value	FC	*P*-value	FC	*P*-value
At2g24762	AtGDU4 (GLUTAMINE DUMPER 4)	-3.1	2.0E-06	-6.7	4.5E-09	-3.2	2.1E-03
At1g80440	Member of the GDU (glutamine dumper)	-2.8	2.6E-06	-6.9	1.3E-09	-14.1	2.7E-07
At5g02760	Protein phosphatase 2C family protein	-2.4	2.7E-02	-11.1	7.9E-06	-2.2	1.4E-04
At5g22920	RING-type Zinc finger protein	-2.3	1.6E-05	-4.8	1.3E-08	-11.6	7.4E-05
At1g12200	zinc finger (C3HC4-type RING finger)	-2.3	6.9E-05	-2.1	2.4E-04	-2.2	8.2E-04
At1g73830	BEE3 (BR ENHANCED EXPRESSION 3)	-2.2	3.9E-03	-2.7	5.3E-04	-5.8	5.7E-05
At2g44130	Galactose oxidase/kelch repeat superfamily	-2.2	1.7E-08	-2.3	8.1E-09	-7.1	7.8E-06
At2g40610	ATEXPA8 (EXPANSIN A8)	-2.1	8.8E-03	-5.4	5.9E-06	-3.8	2.4E-03
At4g30110	HMA2; cadmium-transporting ATPase	-2.1	5.2E-06	-4.7	9.8E-10	-2.2	5.7E-03
At2g15890	MEE14 (maternal effect embryo arrest 14)	-2.1	1.4E-03	-3.0	3.7E-05	-4.0	1.9E-05
At3g46130	MYB111 (MYB DOMAIN PROTEIN 111)	-2.1	1.4E-07	-2.3	2.2E-08	-2.0	3.5E-03
At1g23390	Kelch repeat-containing F-box family protein	-2.0	3.0E-05	-4.9	2.0E-09	-12.3	5.1E-07
At5g51560	Leucine-rich repeat protein kinase family	-2.0	2.6E-05	-2.5	1.2E-06	-2.2	5.3E-03

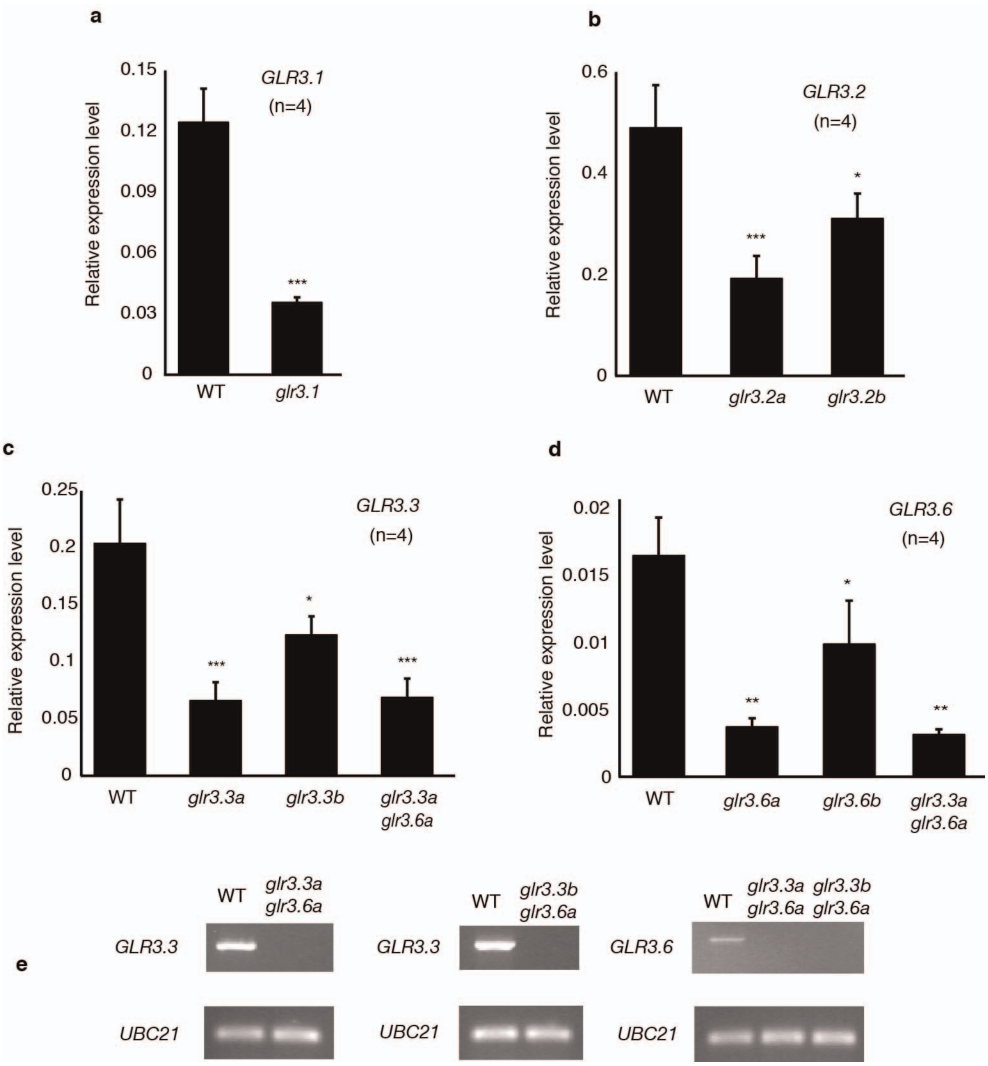
Extended Data Figure 7 | Selected genes for which expression was altered upon current injection. **a**, List of the *JAZ* genes that were upregulated 1 h after current injection (CI) into leaf 8 (this study), in leaf 13 at 1 h after wounding leaf 8 (this study), or in wounded leaves of 18-day-old plants (from ref. 32). **b**, Venn diagram showing downregulated (>twofold, $P < 0.05$) genes for current injected leaf 8 (this study), for leaf 13 from plants wounded on leaf 8 (this study), and for wounded rosette leaves ('rosette after wounding', from ref. 32). **c**, List of common genes that were downregulated more than twofold ($P \leq 0.05$) 1 h after current injection into leaf 8 (this study), in leaf 13 1 h after wounding leaf 8 (leaf 13, this study), and in wounded leaves of 18-day-old plants (ref. 32), FC, fold change.



Extended Data Figure 8 | Effect of inhibitors and *rbohD* on WASP generation and *JAZ10* expression. **a**–**c**, Inhibitors were tested for their effects on WASP generation. **a**, Diphenyleneiodonium chloride (DPI; 50 μM in H_2O containing 1% v/v DMSO), **b**, catalase (100 U μl^{-1} in H_2O) and **c**, lanthanum chloride ($LaCl_3$, 2 mM in H_2O) were infiltrated into leaf 8 at 25–30 min before wounding. After wounding leaf 8, WASP amplitude and duration were measured on leaf 13. For controls leaf 8 was infiltrated only with carrier. *$P < 0.05$ (± s.d.). **d**, *JAZ10* transcript levels in leaf 13 following infiltration of DPI (50 μM in H_2O containing 1% v/v DMSO) into leaf 8 followed 30 min later by wounding leaf 8 (± s.e.m.). Controls (CON) were infiltrated with carrier. **e**, Similar wound-induced expression of *JAZ10* in WT and *rbohD* plants. Plants (wild type or *rbohD-dSpm*) were wounded on leaf 8 (W). After 1 h leaves 8 and 13 were harvested and *JAZ10* expression measured by qRT–PCR (± s.d.). U, unwounded; W, wounded.

Locus	annotation	Stock name	Leaf 8 Amplitude (mV)	Leaf 8 Duration (s)	Leaf 13 Amplitude (mV)	Leaf 13 Duration (s)	n
-	*WT*	Col-0	-76±11	163±30	-51±9	78±20	33
At3g04110	*glr1.1*	salk_057748	-78±12	459±42	-46±5	63±18	6
At3g07520	*glr1.4*	salk_129955	-81±23	149±66	-54±14	81±9	5
At2g24710	*glr2.3*	salk_113260	-78±20	140±12	-32±11	79±8	7
At2g29120	*glr2.7*	salk_121990	-96±12	213±34	-43±10	74±11	5
At2g29110	*glr2.8*	salk_111695	-84±17	272±32	-55±8	92±15	5
At2g29100	*glr2.9*	salk_125496	-59±14	219±29	-43±4	103±23	4
At2g17260	*glr3.1*	salk_063873	-81±18	110±25	-48±13	10±8	9
At4g35290	*glr3.2a*	salk_150710	-87±7	37±6	-46±11	34±4	8
At4g35290	*glr3.2b*	salk_133700	-86±10	99±26	-29±14	33±23	8
At1g42540	*glr3.3a*	salk_099757	-91±11	51±10	-47±10	36±18	9
At1g42540	*glr3.3b*	salk_077608	-71±15	36±6	-40±18	23±5	9
At1g05200	*glr3.4*	salk_079842	-75±16	70±9	-36±10	59±16	7
At2g32390	*glr3.5*	salk_035264	-83±9	94±15	-58±16	63±13	7
At3g51480	*glr3.6a*	salk_091801	-73±12	116±18	-44±6	16±7	9
At3g51480	*glr3.6b*	salk_035353	-65±12	118±10	-42±19	29±16	6
At2g46450	*cngc12*	salk_092622	-69±16	299±31	-36±4	87±14	6
At5g14870	*cngc18*	sail_191_H04	-87±6	430±14	-46±6	74±12	6
At3g17700	*cngc20*	salk_129133	-70±8	>400	-40±10	88±11	4
At3g27170	*clc-b*	salk_027349	-91± 3	309±11	-42±12	55±6	5
At4g35440	*clc-e*	salk_142812	-86±18	148±31	-50±9	48±4	6
At1g04690	*kab1*	salk_056819	-83±15	156±25	-47±10	78±27	7
At2g25600	*akt6*	salk_136050	-84±16	144±24	-52±10	93±25	6
At3g02850	*skor*	salk_097435	-71±16	209±56	-47±12	74±20	7
At5g55630	*tpk1*	salk_146903	-81±11	125±12	-33±12	45±35	8
At4g18160	*tpk3*	salk_049137	-82±14	117±13	-51±12	48±9	9
At5g10220	*annat6*	salk_043207	-81±12	146±21	-39±12	64±21	7
At1g59870	*pen3*	salk_000578	-91±10	147±25	-53±12	80±8	7
At4g11150	*vha-e1*	salk_019365	-81±7	>400	-42±4	81±15	5
At3g01390	*vha-g*	salk_087613	-87±10	232±76	-71±9	87±18	5
At2g21410	*vha-a2*	salk_142642	-79±9	177±17	-53±9	118±19	7
At2g18960	*aha1*	salk_118350	-63±20	109±31	-36±19	70±28	9
At4g30190	*aha2*	salk_073730	-76±19	118±38	-46±9	78±35	8
At5g57350	*aha3*	sail_810_08	71±17	127±38	-58±10	79±19	7
At5g62670	*aha11*	salk_152723	-77±15	101±33	-53±14	87±20	9
At5g47910	*rbohD*	salk_070610	-92±92	98±36	-53±12	61±13	9
At5g47910	*rbohD*	dSpm	-84±20	104±14	-50±23	70±19	7
At2g18960 At5g57350	*aha1aha3*	salk_118350 sail_810_08	-79±13	202±47	-41±11	126±29	7
At4g30190 At5g57350	*aha2aha3*	salk_073730 sail_810_08	-82±10	128±33	-44±15	72±30	8
At1g42540 At3g51480	*glr3.3aglr3.6a*	salk_099757 salk_091801	-55±18	8.5±1.7	0±0	0±0	11
At1g42540 At3g51480	*glr3.3bglr3.6a*	salk_077608 salk_091801	-51±16	9.3±1.3	0±0	0±0	7

Extended Data Figure 9 | Characterization of wound-activated surface potential changes (WASPs) in homozygous T-DNA insertion lines. Leaf 8 was wounded and the surface potential was monitored in leaf 8 and distal leaf 13. For leaf 8, an electrode was placed 3 cm from the leaf apex wound (Fig. 1a, position e3). All measurements for leaf 13 were from electrodes placed on the petiole 1 cm from the centre of the rosette (position e3′ in Extended Data Fig. 2c). *n*, number of experiments. Values are means ± s.d. Mutants displaying WASP durations of <60 s in leaf 8 or <40 s in leaf 13 are highlighted.



Extended Data Figure 10 | Relative expression levels of *GLRs* in the wild type and in T-DNA insertion lines. **a**, Level of *GLR3.1* transcripts in *glr3.1a* (Salk_063873). **b**, Level of *GLR3.2* transcripts in *glr3.2a* (Salk_150710) and *glr3.2b* (Salk_133700). **c**, Level of *GLR3.3* transcripts in *glr3.3a* (Salk_099757), *glr3.3b* (Salk_077608) and double mutant *glr3.3a glr3.6a*. **d**, Level of *GLR3.6* transcripts in *glr3.6a* (Salk_091801), *glr3.6b* (Salk_035353) and double mutant *glr3.3a glr3.6a*. In all cases leaves were harvested from unwounded plant. Significant differences to the wild type are indicated, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ (± s.d.). **e**, RT–PCR analyses of the expression pattern of *GLR3.3* and *GLR3.6* genes in *glr3.3a glr3.6a* and *glr3.3b glr3.6a* double mutants. *UBC21* was the reference transcript.

doi:10.1038/nature12419

An observational correlation between stellar brightness variations and surface gravity

Fabienne A. Bastien[1], Keivan G. Stassun[1,2], Gibor Basri[3] & Joshua Pepper[1,4]

Surface gravity is a basic stellar property, but it is difficult to measure accurately, with typical uncertainties of 25 to 50 per cent if measured spectroscopically[1,2] and 90 to 150 per cent if measured photometrically[3]. Asteroseismology measures gravity with an uncertainty of about 2 per cent but is restricted to relatively small samples of bright stars, most of which are giants[4–6]. The availability of high-precision measurements of brightness variations for more than 150,000 stars[7,8] provides an opportunity to investigate whether the variations can be used to determine surface gravities. The Fourier power of granulation on a star's surface correlates physically with surface gravity[9,10]: if brightness variations on timescales of hours arise from granulation[11], then such variations should correlate with surface gravity. Here we report an analysis of archival data that reveals an observational correlation between surface gravity and root mean squared brightness variations on timescales of less than eight hours for stars with temperatures of 4,500 to 6,750 kelvin, log surface gravities of 2.5 to 4.5 (cgs units) and overall brightness variations of less than three parts per thousand. A straightforward observation of optical brightness variations therefore allows a determination of the surface gravity with a precision of better than 25 per cent for inactive Sun-like stars at main-sequence to giant stages of evolution.

Brightness variations of Sun-like stars are driven by many factors, including granulation[12], oscillations[11], rotation and magnetic activity[13]. As they evolve from high-surface-gravity (high-g) dwarfs to low-g giants, their convective zones deepen, they rotate more slowly, their magnetic activity diminishes, and their oscillation and granulation timescales increase, all of which change the nature of the brightness variations. It has been previously demonstrated that the power in granulation (as traced by the Fourier spectrum of the brightness variations) is inversely proportional to ν_{max}, the peak frequency of Sun-like acoustic oscillations[9,12]. Given that ν_{max} is itself proportional to g (ref. 11), it follows that g should manifest in brightness variations on timescales that trace granulation. Although physically we expect this, it is not immediately apparent that brightness variations can be used as an effective determinant of g because other phenomena not directly related to g—most importantly spots, plage and other sources of brightness variations driven by the star's magnetic activity—probably dominate the observed brightness variations. It is therefore necessary to filter out the brightness variations arising from these phenomena, which occur on timescales of hours to days, while preserving the brightness variations related to granulation and g on timescales of minutes to hours.

Using long-cadence (30 min) light curves from Quarter 9 of NASA's Kepler Mission[14], and representing them using the Filtergraph data visualization tool[15], we observe clear patterns in the evolutionary properties of stars encoded in three simple measures of their brightness variations[8] (Fig. 1): range (R_{var}), number of zero crossings (X_0), and root mean square on timescales shorter than 8 h (to which we will refer as '8-hr flicker' or F_8). Relating these measures to g determined asteroseismically from a sample of Kepler stars[4], we find distinctive features that highlight the way stars evolve in this three-dimensional space, making up an evolutionary diagram of photometric variability. Within this diagram, we find a vertical cloud of points, largely made up of high-g dwarfs, that have large R_{var}, small X_0 and low F_8 values. We observe a tight sequence of stars—a 'flicker-floor' sequence that defines a prominently protruding lower envelope in R_{var}—spanning gravities from dwarfs to giants. Sun-like stars of all evolutionary states evidently move onto this sequence only when they have a large X_0, which in turn implies low stellar activity.

We find that g is encoded in F_8, yielding a tight correlation between the two (Fig. 2). Moreover, using 11 yr of SOHO Virgo[16,17] light curves of the Sun and sampling them at the same cadence as the Kepler long-cadence light curves, we find that the Sun's (constant) g is also measurable using F_8, which remains invariant throughout the 11-yr solar activity cycle even while the Sun's R_{var} and X_0 values change considerably from the spot-dominated solar maximum to the nearly spotless solar minimum (Fig. 1). From the Sun's behaviour, we infer that a large portion of the Kepler stars' vertical scatter within the vertical cloud at the left of the diagram (Fig. 1) may be driven by solar-type cyclic activity variations. Most importantly, the Sun's true g fits our empirical relation (Fig. 2), and the g value of any Sun-like Kepler star from dwarf to giant may be inferred from this relation with an accuracy of 0.06–0.10 dex (Supplementary Information).

Asteroseismic analyses derive g from the properties of stellar acoustic oscillations[4,18–20]. Given that near-surface convection drives both these oscillations and granulation, and given the brightness variability timescales to which F_8 is sensitive, we suggest that a combination of different types of granulation (with typical solar timescales ranging from ~30 min to ~30 h; ref. 21) drives the manifestation of g in this metric. The precise timescales of these phenomena in solar-type stars depend strongly on the stellar evolutionary state and, hence, also on g (refs 5,9,10,22). Acoustic oscillations, whose amplitudes are sensitive to g (ref. 5), may provide an increasingly important contribution to F_8 as stars evolve into subgiants and giants and the amplitudes and timescales of these oscillations increase[5,9,10]. At some point, the pressure-mode and granulation timescales cross[9], which may lead to a breakdown of our F_8–g relation at very low values of g.

By using F_8 to measure g, we can construct a photometric variability evolutionary diagram for most stars observed by Kepler, even for stars well beyond the reach of asteroseismic and spectroscopic analysis (Fig. 3). By coding this diagram according to stellar temperature and rotation period, we may trace the physical evolution of Sun-like stars as follows: stars begin as main-sequence dwarfs with large photometric R_{var} values and small X_0 values, presumably driven by simple rotational modulation of spots at relatively short rotation periods. As the stars spin down to longer rotation periods, their brightness variations first become steadily 'quieter' (systematically lower R_{var}) but then become suddenly and substantially more complex (larger X_0) as they reach the flicker floor. Some stars reach the floor only after beginning their evolution as low-g subgiants, having moved to the right (higher F_8) as their effective temperatures begin rapidly dropping. Other stars

[1]Department of Physics and Astronomy, Vanderbilt University, 1807 Station B, Nashville, Tennessee 37235, USA. [2]Physics Department, Fisk University, 1000 17th Avenue North, Nashville, Tennessee 37208, USA. [3]Astronomy Department, University of California, Hearst Field Annex, Berkeley, California 94720, USA. [4]Physics Department, Lehigh University, 27 Memorial Drive West, Bethlehem, Pennsylvania 18015, USA.

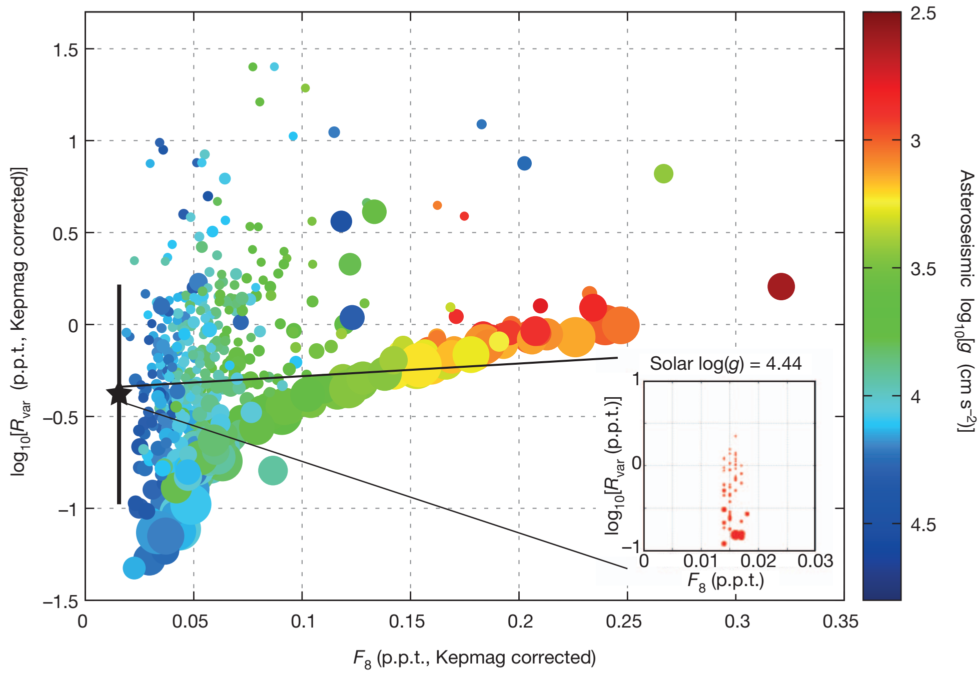


Figure 1 | Simple measures of brightness variations reveal a fundamental 'flicker sequence' of stellar evolution. We establish the evolutionary states of stars with three simple measures of brightness variations[8]. The abscissa, 8-h flicker (F_8), measures brightness variations on time scales of 8 h or less. The ordinate, R_{var}, yields the largest amplitude of the photometric variations in a 90-d timeframe. The number of zero crossings, X_0 (symbol size; ranging from 0.01 to 2.1 crossings per day), conveys the large-scale complexity of the light curve. We correct both R_{var} and F_8 for their dependence on Kepler magnitude (Kepmag). Colour represents asteroseismically determined g. We observe two populations of stars: a vertical cloud composed of high-g dwarfs and some subgiants, and a tight sequence—the flicker floor—spanning a range of g from dwarfs to giants. The typically large R_{var} values of stars in the cloud, coupled with their simpler light curves (small X_0), implies brightness variations driven by rotational modulation of spots. In contrast, large X_0 values characterize stars on the sequence. The F_8 values of stars in this sequence increase inversely with g because the physical source of F_8 is sensitive to g. R_{var} also increases with F_8 along the floor, because F_8 is a primary contributor to R_{var} (as opposed to starspots above the floor). Stars with a given F_8 value cannot have R_{var} less than that implied by F_8 itself: quiet stars accumulate on the flicker floor because they are prevented from going below it by the statistical definition of the two quantities. Stars above the floor have larger amplitude variations on longer timescales that set R_{var}. The large star symbol with vertical bars and the inset show the Sun's behaviour over the course of its 11-yr magnetic cycle. The Sun's F_8 value is largely invariant over the course of its cycle, just as its g value is invariant. p.p.t., parts per thousand.

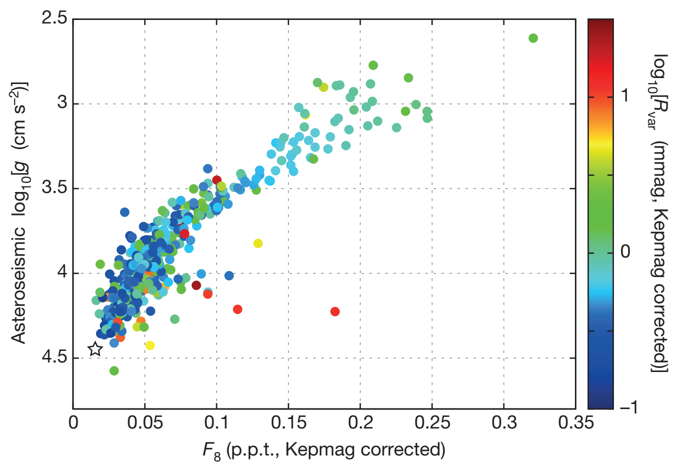


Figure 2 | Stellar surface gravity manifests in a simple measure of brightness variations. The same stars from Fig. 1 with Kepler Quarter 9 data. Asteroseismically determined[4] g shows a tight correlation with F_8. Colour represents the R_{var} of the stars' brightness variations; outliers tend to have large brightness variations. Excluding these outliers, a cubic-polynomial fit through the Kepler stars and through the Sun (large star symbol) shows a median absolute deviation of 0.06 dex and a root mean squared deviation of 0.10 dex (Supplementary Information). To simulate how the solar g would appear in the archival data we use to measure g for other stars, we divide the solar data into 90-d 'quarters'. Our F_8–g relation measured over multiple quarters then yields a median solar g of 4.442 with a median absolute deviation of 0.005 dex and a root mean squared error of 0.009 dex (the true solar g is 4.438).

join the sequence while still dwarfs; these are easily identified in our diagram by their drastically increased X_0 values at very low F_8. Evidently some dwarf stars become magnetically quiet while still firmly on the main sequence, whereas others do not reach the floor until they begin to swell considerably. We note that the Sun seems to approach the flicker floor at solar minimum; its R_{var} value becomes quite low and its X_0 value strongly increases (Fig. 1).

A star's main-sequence mass and initial spin probably determine where along the flicker-floor sequence it ultimately arrives, because the slope of a star's trajectory in our diagram is essentially determined by the ratio of its spin-down timescale (downward motion) and structural evolutionary timescale (rightward motion). Regardless, once on the floor all stars evolve along this sequence and stay on it as they move up to the red-giant branch, their effective temperatures steadily dropping as their surfaces rapidly expand. Despite their very slow rotation as subgiants and giants on the flicker-floor sequence, their photometric R_{var} is steadily driven upwards by the increasing F_8, which reflects the stars' continually decreasing g. The increasing R_{var} and F_8 values of subgiants and giants on the flicker floor is probably the result of the increasingly important contribution of radial and non-radial pulsations to the overall brightness variations[23,24].

A few stars appear as outliers to the basic picture we have presented here; these are seen towards the right of the vertical cloud of points in our evolutionary diagram (Fig. 3). Some active dwarfs have higher F_8 values than expected for their g values. Frequent strong flares can boost F_8 as currently defined, and some hotter dwarfs are pulsators with enough power near 8 h to increase their F_8 values. A few such cases appear also in the asteroseismic sample (Fig. 1). Some lower-g stars have R_{var} values above the flicker floor owing to the presence of magnetic

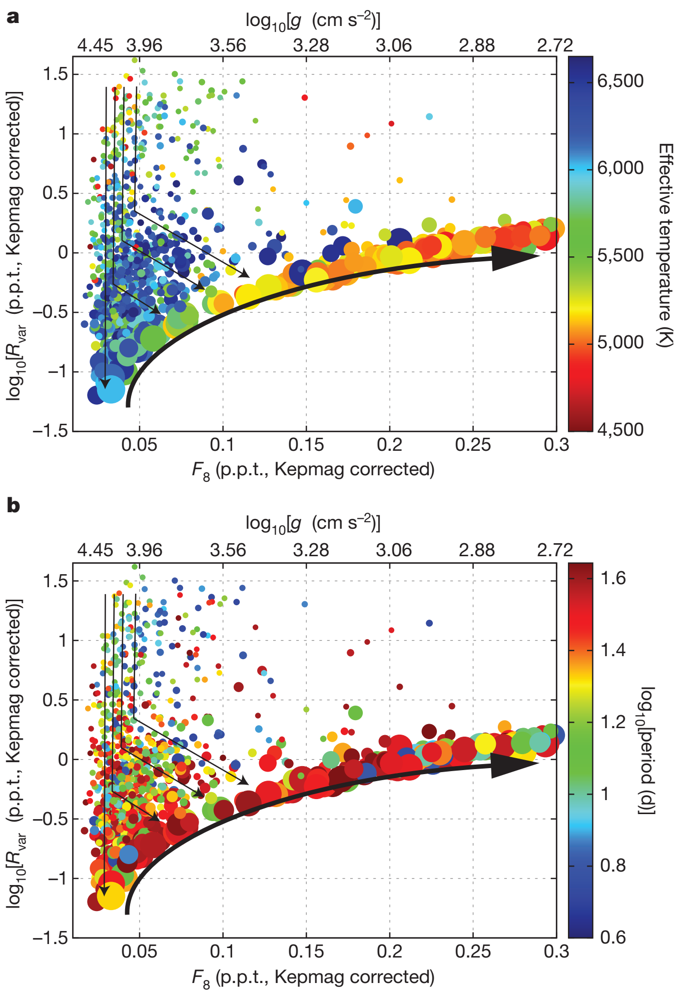


Figure 3 | An integrative view of stellar evolution in a new diagram of brightness variations. Same as Fig. 1, but for Kepler stars lacking asteroseismic g. We include a g scale at the top (from conversion of the F_8 scale at bottom via our calibrated relationship). Here we selected stars with Kepler magnitudes between 11.0 and 11.85 to limit the sample to ~1,000 stars for visual clarity (1,012 points are shown). We removed objects that are potentially blended (Kepler flux contamination greater than 0.05) as well as those that may be galaxies (Kepler star/galaxy flag other than 0). Arrows schematically indicate the evolutionary paths of Sun-like stars in this diagram. Stars generally move from top to bottom, as the overall brightness fluctuations due to spots decrease with time, and then from left to right as their g values decrease. All stars eventually arrive on the flicker-floor sequence and evolve along it. **a**, Colour represents effective temperature. Stars cool as they evolve from left to right, from dwarfs to red giants. We restricted the effective temperatures to be 4,500–6,650 K, using the revised temperature scale for Kepler stars[30]. **b**, Same as **a**, but colour-coded by the dominant periodicity in the light curve. We limited the sample to stars with dominant periods longer than 3 d (to eliminate very rapidly rotating active stars) and shorter than 45 d (half the Kepler 90-d data interval). This period traces rotation for unevolved stars and pulsations for evolved ones. Dwarfs generally show the expected spin-down sequence with decreasing R_{var} (correlated with the level of surface magnetic activity). Subgiants and giants broadly display very slow rotation, as expected.

activity[25], slow radial pulsations or secular drifts. Finally, a few outliers are simply due to data anomalies. As our technique is refined, these exceptions should be treated carefully before assigning a F_8-based g value, particularly for high-F_8 stars for which R_{var} is greater than ~3 parts per thousand. They constitute a small fraction of the bulk sample, and most of them can be identified as one of the above cases.

Common to all of the stars along the flicker floor is the virtual absence of spot activity as compared with their higher-R_{var} counterparts; short-timescale phenomena such as granulation and oscillations dominate the brightness variations. Given that spots probably suppress acoustic oscillations in the Sun and other dwarf stars[5,26–28], the large X_0 values of stars along this sequence may partly reflect the ability of short-timescale processes to manifest more strongly now that large spots no longer impede them, along with the increasing complexity of the convective variations. As the stars evolve into full-fledged red giants and beyond, the principal periodicity in their brightness variations increasingly reflects shorter-period oscillations, as opposed to their inherently long-period rotation, because oscillations become dominant over magnetic spots.

It may be possible to differentiate between stars with similar g values but different internal structures (for example first-ascent red giants versus helium-burning giants) through application of a sliding timescale of F_8 as a function of g, where the sliding timescale would capture the changing physical granulation timescales with evolutionary state[20]. Moreover, the behaviour of stars on the flicker floor may explain the source of radial velocity 'jitter' that now hampers planet detection through radial velocity measurements[29].

Received 2 April; accepted 20 June 2013.

1. Valenti, J. & Fischer, D. A. Spectroscopic properties of cool stars (SPOCS). I. 1040 F, G, and K dwarfs from Keck, Lick, and AAT planet search programs. *Astrophys. J.* **159,** 141–166 (2005).
2. Ghezzi, L. *et al.* Stellar parameters and metallicities of stars hosting Jovian and Neptunian mass planets: a possible dependence of planetary mass on metallicity. *Astrophys. J.* **720,** 1290–1302 (2010).
3. Brown, T. M., Latham, D. W., Everett, M. E. & Esquerto, G. A. Kepler Input Catalog: photometric calibration and stellar classification. *Astron. J.* **142,** 112–129 (2011).
4. Chaplin, W. J. *et al.* Ensemble asteroseismology of solar-type stars with the NASA Kepler mission. *Science* **332,** 213–216 (2011).
5. Huber, D. *et al.* Testing scaling relations for solar-like oscillations from the main sequence to red giants using Kepler data. *Astrophys. J.* **743,** 143–152 (2011).
6. Stello, D. *et al.* Asteroseismic classification of stellar populations among 13,000 red giants observed by Kepler. *Astrophys. J.* **765,** L41–L45 (2013).
7. Basri, G. *et al.* Photometric variability in Kepler target stars: the Sun among stars – a first look. *Astrophys. J.* **713,** L155–L159 (2010).
8. Basri, G. *et al.* Photometric variability in Kepler target stars. II. An overview of amplitude, periodicity, and rotation in the First Quarter data. *Astron. J.* **141,** 20–27 (2011).
9. Mathur, S. *et al.* Granulation in red giants: observations by the Kepler mission and three-dimensional convection simulations. *Astrophys. J.* **741,** 119–130 (2011).
10. Kjeldsen, H. & Bedding, T. R. Amplitudes of solar-like oscillations: a new scaling relation. *Astron. Astrophys.* **529,** L8–L11 (2011).
11. Brown, T. M., Gilliland, R. L., Noyes, R. W. & Ramsey, L. W. Detection of possible p-mode oscillations on Procyon. *Astrophys. J.* **368,** 599–609 (1991).
12. Gilliland, R. L. *et al.* Kepler mission stellar and instrument noise properties. *Astrophys. J.* **197** (suppl.), 6–24 (2011).
13. Strassmeier, K. G. Starspots. *Astron. Astrophys. Rev.* **17,** 251–308 (2009).
14. Borucki, W. J. *et al.* Kepler planet-detection mission: introduction and first results. *Science* **327,** 977–980 (2010).
15. Burger, D. *et al.* An interactive web application for visualization of astronomy datasets. *Astron. Comput.* (in the press); preprint at http://arxiv.org/abs/1307.4000 (2013).
16. Fröhlich, C. *et al.* First results from VIRGO, the experiment for helioseismology and solar irradiance monitoring on SOHO. *Sol. Phys.* **170,** 1–25 (1997).
17. Basri, G., Walkowicz, L. M. & Reiners, A. Comparison of Kepler photometric variability with the Sun on different timescales. *Astrophys. J.* **769,** 37–49 (2013).
18. Brown, T. M. & Gilliland, R. L. Asteroseismology. *Annu. Rev. Astron. Astrophys.* **32,** 37–82 (1994).
19. Christensen-Dalsgaard, J. Physics of solar-like oscillations. *Sol. Phys.* **220,** 137–168 (2004).
20. Chaplin, W. J. & Miglio, A. Asteroseismology of solar-type and red giant stars. *Annu. Rev. Astron. Astrophys.* (in the press).
21. Dumusque, X., Udry, S., Lovis, C., Santos, N. C. & Monteiro, M. J. P. F. G. Planetary detection limits taking into account stellar noise. I. Observational strategies to reduce stellar oscillation and granulation effects. *Astron. Astrophys.* **525,** 140–151 (2011).
22. Kjeldsen, H. & Bedding, T. R. Amplitudes of stellar oscillations: the implications for asteroseismology. *Astron. Astrophys.* **293,** 87–106 (1995).
23. Henry, G. W., Fekel, F. C., Henry, S. M. & Hall, D. S. Photometric variability in a sample of 187 G and K giants. *Astrophys. J.* **130** (suppl.), 201–225 (2000).
24. Gilliland, R. L. Photometric oscillations of low-luminosity red giant stars. *Astron. J.* **136,** 566–579 (2008).
25. Schröder, C., Reiners, A. & Schmitt, J. H. M. M. Ca II HK emission in rapidly rotating stars. Evidence for an onset of the solar-type dynamo. *Astron. Astrophys.* **493,** 1099–1107 (2009).
26. Chaplin, W. J., Elsworth, Y., Isaak, G. R., Miller, B. A. & New, R. Variations in the excitation and damping of low-l solar p modes over the solar activity cycle. *Mon. Not. R. Astron. Soc.* **313,** 32–42 (2000).

27. Komm, R. W., Howe, R. & Hill, F. Solar-cycle changes in GONG p-mode widths and amplitudes 1995–1998. *Astrophys. J.* **531,** 1094–1108 (2000).
28. Chaplin, W. J. *et al.* Evidence for the impact of stellar activity on the detectability of solar-like oscillations observed by Kepler. *Astrophys. J.* **732,** L5–L10 (2011).
29. Bastien, F. A. *et al.* Radial velocity variations of photometrically quiet, chromospherically inactive Kepler stars: a link between RV jitter and photometric flicker. *Astron. J.* (submitted).
30. Pinsonneault, M. *et al.* A revised effective temperature scale for the Kepler Input Catalog. *Astrophys. J.* **199** (suppl.), 30–51 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements The research described in this paper makes use of Filtergraph (http://filtergraph.vanderbilt.edu), an online data visualization tool developed at Vanderbilt University through the Vanderbilt Initiative in Data-intensive Astrophysics. We acknowledge discussions with P. Cargile, K. Carpenter, W. Chaplin, D. Huber, M. Paegert, M. Sinha and D. Weintraub. We thank D. Huber and T. Metcalfe for sharing the average asteroseismic parameters of Kepler stars with us. F.A.B. acknowledges support from a NASA Harriet Jenkins Fellowship and a Vanderbilt Provost Graduate Fellowship. F.A.B. and K.G.S. acknowledge NSF PAARE grant AST-0849736.

Author Contributions F.A.B. and K.G.S. contributed equally to the identification and analysis of the major correlations. F.A.B. principally wrote the first version of the manuscript. K.G.S. prepared the figures. G.B. calculated the variability statistics of the Kepler light curves and performed an independent check of the analysis. J.P. checked against biases in the datasets. All authors contributed to the interpretation of the results and to the final manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.A.B. (fabienne.a.bastien@vanderbilt.edu).

doi:10.1038/nature12425

Measurement of a solid-state triple point at the metal–insulator transition in VO_2

Jae Hyung Park[1], Jim M. Coy[1], T. Serkan Kasirga[1], Chunming Huang[1], Zaiyao Fei[1], Scott Hunter[1] & David H. Cobden[1]

First-order phase transitions in solids are notoriously challenging to study. The combination of change in unit cell shape, long range of elastic distortion and flow of latent heat leads to large energy barriers resulting in domain structure, hysteresis and cracking. The situation is worse near a triple point, where more than two phases are involved. The well-known metal–insulator transition in vanadium dioxide[1], a popular candidate for ultrafast optical and electrical switching applications, is a case in point. Even though VO_2 is one of the simplest strongly correlated materials, experimental difficulties posed by the first-order nature of the metal–insulator transition as well as the involvement of at least two competing insulating phases have led to persistent controversy about its nature[1–4]. Here we show that studying single-crystal VO_2 nanobeams[5–16] in a purpose-built nanomechanical strain apparatus allows investigation of this prototypical phase transition with unprecedented control and precision. Our results include the striking finding that the triple point of the metallic phase and two insulating phases is at the transition temperature, $T_{tr} = T_c$, which we determine to be 65.0 ± 0.1 °C. The findings have profound implications for the mechanism of the metal–insulator transition in VO_2, but they also demonstrate the importance of this approach for mastering phase transitions in many other strongly correlated materials, such as manganites[17] and iron-based superconductors[18].

The metal–insulator transition (MIT) in VO_2 is accompanied by a large and rapid change in the conductivity and optical properties, with potential uses in switching and sensing. VO_2 has recently received renewed attention as a convenient strongly correlated material for the application of new ultrafast[19–21] and microscopy[22,23] techniques, ionic gating[24] and improved computational approaches[3,4]. However, the problems associated with bulk or film samples that consist of a complex of multiple solid phases and domains under highly non-uniform strain, as well as compositional variations such as oxygen vacancies[25] and hydrogen doping[26], make it almost impossible to disentangle the underlying parameters on which rigorous understanding can be built. The experiments described here eliminate these problems, allowing unprecedented control of the MIT and accurate determination of the underlying phase stability diagram of pure VO_2.

Figure 1a illustrates the structures of the phases involved in the MIT. In every phase there are two interpenetrating sets of parallel chains of vanadium atoms each surrounded by six oxygen atoms forming a distorted octahedron (the oxygen atoms are not shown). In the high-temperature metallic (rutile, R) phase all the chains are straight and periodic, whereas in the low-temperature insulating (monoclinic M1) phase every chain is dimerized. There are also two other known insulating phases: monoclinic M2, in which only one set of chains is dimerized, and triclinic T, which is intermediate between M1 and M2. The existence of both M1 and M2, with similar dielectric properties yet different magnetic properties, provides constraints on the theory of the MIT; for example, it rules out a purely Peierls-type mechanism[2]. In the older literature the MIT is taken to occur between R and M1, although recent studies[8–10,23] have shown that M2 domains occur in most VO_2 samples near the MIT, raising the question of its role in the transition.

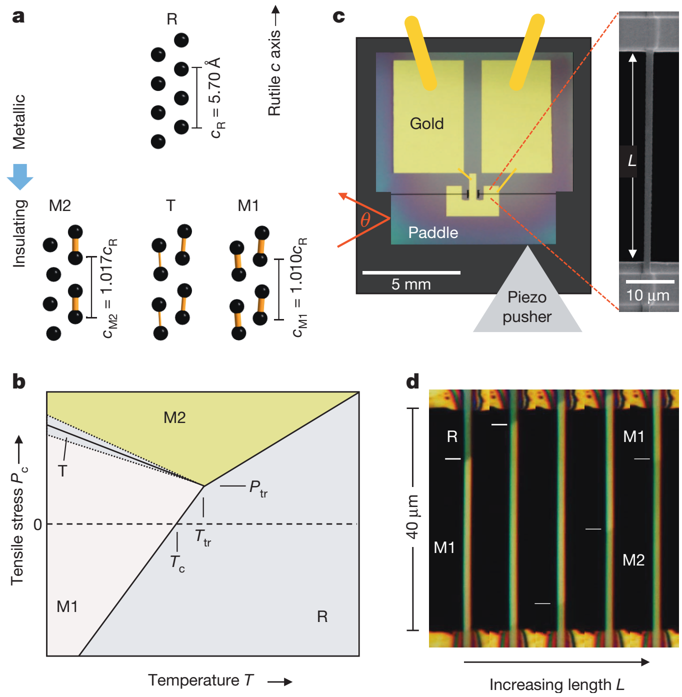


Figure 1 | Control of the metal–insulator transition in VO_2 using uniaxial stress. a, Arrangement of vanadium ions in the phases involved in the MIT, indicating their different vanadium chain periods and dimerization (yellow). **b**, Expected layout of the stress–temperature phase diagram near the MIT, showing the transition temperature T_c at zero stress. **c**, Experimental geometry, showing an electron micrograph (right) of a VO_2 nanobeam suspended across a slot of width L in a silicon chip (left, optical micrograph) whose width is controlled by pushing on the paddle and measured by deflection of a laser beam. The yellow lines signify gold wire bonds. **d**, Series of optical images showing movement of the R–M1 and M1–M2 interfaces as L is increased in roughly 100-nm steps at 64 °C (device P7, 40 µm gap).

The largest difference in unit cell shape between R, M1 and M2 is along the pseudo-rutile c axis (the vanadium chain axis), with $c_R = 5.700$ Å, $c_{M1} = 5.755$ Å and $c_{M2} = 5.797$ Å, as indicated in Fig. 1a. Compressive strain along this axis in an epitaxial film can lower the transition to room temperature[15,25]; thus, applying uniaxial tensile stress P_c along it can be used to control the transition[13,15]. A stability diagram in the P_c–T plane (with all other stress components zero) is expected to have the layout indicated in Fig. 1b. A shaded region indicates where the T phase occurs[7,27]. The effect of P_c on the phase stability (Fig. 1b) resembles that of stress along the $[110]_R$ axis[27] and of doping[28] by chromium. Rough ideas of the locations of the three phase boundaries have been obtained by modelling bent nanobeams[16]. The triple point (T_{tr}, P_{tr}) has not been located, although M1 and M2 are known to be very close in free energy near the transition[27]. The stress P_{tr} is normally taken to be positive, implying that a perfect unstrained crystal

[1]Department of Physics, University of Washington, Seattle, Washington 98195, USA.

shows a direct transition from M1 to R at T_c. We find that this is not in fact true, and T_{tr} is identical to T_c to within $\pm 0.05\,^{\circ}$C, or one part in 10^4 in absolute temperature. We further determine T_c to be $65.0 \pm 0.1\,^{\circ}$C. In addition we present evidence that in the neighbourhood of T_c the M1 phase can distort continuously under tension into the metastable T phase. These discoveries have deep implications for the physics of the MIT, for the interpretation of many measurements on VO_2 crystals and films, and for mastering the transition with a view to applications.

Our investigations of the MIT rest on the ability to precisely control the length of a suspended single-domain nanobeam and thereby to apply pure uniaxial stress along it, a situation that cannot be achieved in larger crystals because of domain structure. The elements of the experiment are illustrated in Fig. 1c (see Methods). A VO_2 nanobeam is fixed, in some cases with electrical contacts, across a micromachined slot in a silicon chip whose width L can be varied with nanometre precision. We perform measurements only when the nanobeams are straight, so the maximum compressive stress is limited by buckling. By varying L and T, the three phases R, M1 and M2 can be induced and can be differentiated by reflection contrast with linearly polarized light[10], as illustrated in Fig. 1d, as well as by Raman spectroscopy[14] and measurements of electrical resistance. Linearly polarized light also reveals twinning[11], allowing us to select devices in which twinning is absent.

According to the phase diagram in Fig. 1b the state of the nanobeam as a function of L and T should include regions of two-phase coexistence as sketched in Fig. 2a. We find that the suspended part of the nanobeam can indeed be brought into coexistence between any pair of the three phases. The position of the interface changes smoothly and reproducibly with both L and T in between sudden reconfigurations. For the case of M2 + R coexistence we define the interface position y_{M2R} as the shift relative to an initial position such that it increases as R converts to M2. We define y_{M2M1} and y_{M1R} similarly.

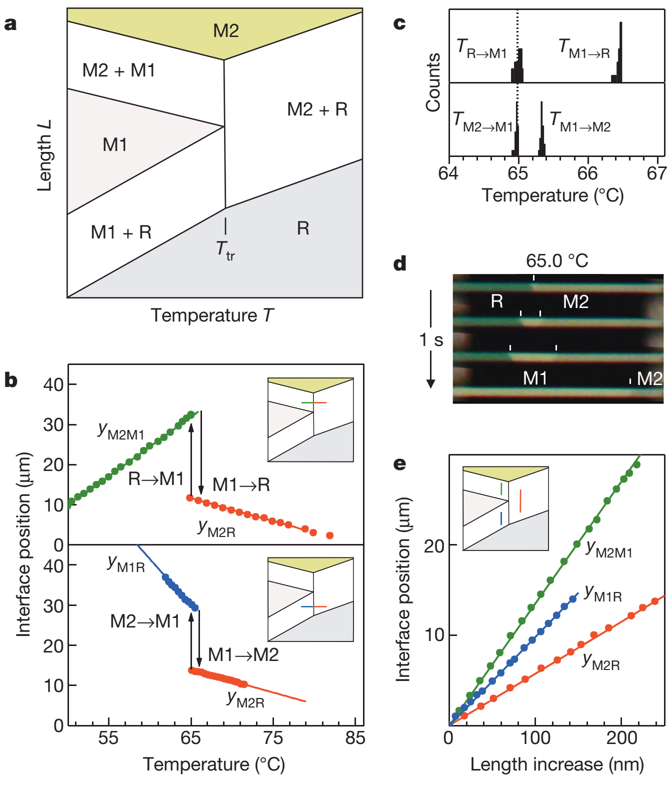


Figure 2 | Temperature and length dependence in coexistence. **a**, Expected configuration of a nanobeam as a function of T and L. **b**, Variation of interface positions with T at fixed L corresponding to moving along the lines in the insets (upper: device P11, 40 μm gap; lower: P9, 20 μm gap). Each interface type is indicated by a colour. **c**, Histograms of temperatures at which reconfigurations occur for 20 cycles sweeping at $0.1\,^{\circ}$C min^{-1} (device P14, 40 μm). **d**, Sequence of images during reconfiguration from M2 + R to M2 + M1 in a nanobeam at the triple point, $65.0\,^{\circ}$C (device P8B, 20 μm). **e**, Variation of interface positions with L at fixed T, corresponding to moving along the vertical lines in the inset (device P14). The fractional differences in lattice constants, α_{ij}, are the inverse slopes of these lines.

The MIT in VO_2 is usually studied as a function of T, without paying close attention to strain or to interconversion between M1 and M2. In undoped samples it is seen in the range 65–68 °C, with a hysteresis of several degrees Celsius, and the value of T_c is not known more precisely than this. In our experiments on nanobeams, as T is varied at fixed L we see the behaviour shown in Fig. 2b, which can be understood with reference to the colour-coded lines in the inset phase diagrams. If we start in M2 + M1 coexistence (Fig. 2b, upper panel, green) and increase T, the interface position y_{M2M1} first moves smoothly as the stress required for phase equilibrium changes[13]. Then at a temperature $T_{M1\to R}$ there is a sudden reconfiguration to M2 + R coexistence (Fig. 2b, upper panel, red) after which the interface position y_{M2R} moves smoothly again. On cooling, the reverse reconfiguration occurs at temperature $T_{R\to M1}$. Starting instead at a smaller length, in M1 + R coexistence (Fig. 2b, lower panel, blue), a jump to M2 + R coexistence (again red) occurs at $T_{M1\to M2}$, whereas the reverse occurs at $T_{M2\to M1}$. Histograms of the reconfiguration temperatures on repeated cycling at $0.1\,^{\circ}$C min^{-1} are shown in Fig. 2c. For this device $T_{M1\to R}$ and $T_{M1\to M2}$ are narrowly peaked at 66.4 and 65.3 °C, respectively; for other devices different values are found. This can be explained by superheating of M1, which varies between devices because the ease of nucleation of the high-temperature phase (R or M2) depends on microscopic details.

In contrast, $T_{R\to M1}$ and $T_{M2\to M1}$ are both peaked at the same temperature, 65.0 °C, indicated by the dotted line in Fig. 2c. In several nanobeams of different sizes, grown on different occasions, these two temperatures always lay in the narrow range between 64.9 and 65.2 °C; moreover, neither storage in air for 6 months nor heating to 200 °C for 1 h changed them, indicating that effects of oxygen vacancies[25] and hydrogen doping[26] were minimal. This observation can be explained as follows. A small amount of M1 is often visible at the interface in M2 + R coexistence, probably because it reduces the elastic energy. On cooling there is therefore no need for nucleation of M1, and reconfiguration occurs as soon as the triple point is reached. In fact, the dynamics of this process can sometimes be observed. Figure 2d shows a sequence of images taken in less than a second during the reconfiguration of a nanobeam after bringing it slowly down to 65.0 °C in M2 + R coexistence. A small pre-existing wedge of M1 at the M2 + R interface rapidly expands to replace the R part of the nanobeam completely. All the above observations thus suggest that the triple point is between 64.9 and 65.2 °C.

We now consider varying L at fixed T. First, in coexistence between any pair of phases the interface position is linear in L, as shown in Fig. 2e. This follows from the fact that the interface moves so as to maintain P_c at the phase equilibrium value. A length increase δL causes an interface shift δy_{M1R}, which changes the natural length by δL to keep the strain constant. This implies that $\delta L = \alpha_{M1R}\delta y_{M1R}$, where $\alpha_{M1R} \equiv c_{M1}/c_R - 1$. Hence y_{M1R} should vary according to $dL/dy_{M1R} = \alpha_{M1R}$, and similarly $dL/dy_{M2M1} = \alpha_{M2M1} \equiv c_{M2}/c_{M1} - 1$ and $dL/dy_{M2R} = \alpha_{M2R} \equiv c_{M2}/c_R - 1 \approx \alpha_{M2M1} + \alpha_{M1R}$. Best linear fits to the data shown give $\alpha_{M2M1} = 0.0074$, $\alpha_{M1R} = 0.0100$ and $\alpha_{M2R} = 0.0174$, close to the values of 0.0073, 0.0098 and 0.0172 calculated from the known lattice constants[28,29].

The ability to control L allows us to confirm the temperature of the triple point and to determine the behaviour very close to it. We exploit the fact that the electrical resistance of the nanobeam, R_n, is sensitive to the phase composition because each phase has a different resistivity[12,13]. The measurements in Fig. 3 are for a device (P10) with indium contacts. Figure 3a shows that at 65.3 °C R_n changes smoothly with L, as a result of a smoothly changing M2 + R interface position for $T > T_{tr}$ (see inset, red line). In contrast, at 63.9 °C it changes in a more complicated way, reflecting the sequence M1 + R→M1→M2 + M1

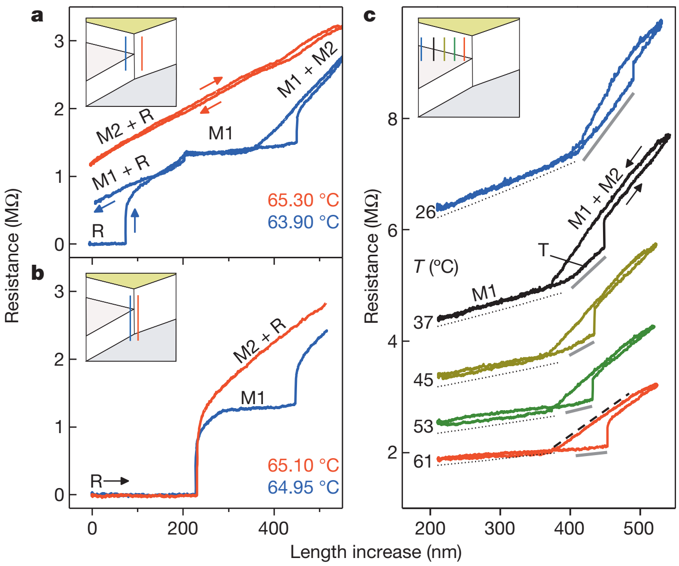


Figure 3 | Resistance–length measurements. **a**, At 65.3 °C, above the triple point (red line in inset), the resistance varies steadily in M2 + R coexistence. At 63.9 °C, below the triple point (blue line in inset), it reflects a sequence of transitions (device P10; L is varied at 8 nm s^{-1}). **b**, Starting in the R state, M2 nucleates if $T \geq 65.10$ °C (red), whereas M1 nucleates if $T \leq 64.95$ °C (blue), implying that these lie on either side of T_{tr} (see inset). **c**, The variation of the resistance with L and T is due to a strain-dependent activation energy in the M1 phase (dotted lines, offset by −0.15 MΩ for clarity) and to conversion of M1 to M2 in coexistence (dashed line). Grey lines indicate an additional resistance rise attributed to the T phase.

expected for $T < T_{tr}$ (see Fig. 3a inset, blue line). Jumps and hysteresis here show that M1 and M2 both require nucleation, which is consistent with the transitions being first order. To establish T_{tr} we measured R_n at a series of closely spaced temperatures, each time preparing the nanobeam in a fully metallic R state by cooling at sufficiently small L for R to be stabilized by compression, and then increasing L until an insulating domain nucleated. At 64.95 °C and below, the domain that appeared was always M1, whereas at 65.10 °C and above it was always M2, implying that T_{tr} was between these two values (see Fig. 3b). This is perfectly consistent with the range of T_{tr} deduced above from the T-sweeping measurements. Including uncertainties from variation between samples, temperature fluctuations and calibration, we conclude that $T_{tr} = 65.0 \pm 0.1$ °C.

Measurements of resistance versus length also yield other useful information, as illustrated in Fig. 3c (see Supplementary Information for details). First, the variation of the resistance of the M1 state with L and T is explained by a linear increase in the activation energy of the resistance with tensile strain $\eta = (L - L_0)/L_0$, L_0 being the effective natural length. The dotted lines are plots of $R_n \propto \exp[-(\Delta_0 + \gamma\eta)/k_B T]$ using coefficient values $\Delta_0 = 0.31$ eV and $\gamma = 0.77$ eV (the uncertainty in γ is 10%), where k_B is Boltzmann's constant. Second, from the variation of R_n in M1 + M2 coexistence (such as that indicated by the dashed line) we can deduce that $\rho_{M2}/\rho_{M1} = 2.3 \pm 0.2$ and that the activation energies of M1 and M2 are the same to within a few per cent. Third, a distinct additional increase in R_n, indicated by the solid grey lines, precedes the nucleation of M2 from M1. This can be explained by a continuous distortion of M1 into the T phase, which we immediately infer has a higher resistivity than M1 and is unstable relative to M2 at all temperatures from T_{tr} to below 26 °C.

Although we cannot measure the axial stress P_c directly, we can realize the condition $P_c = 0$ simply by breaking a nanobeam with a micromanipulator after other measurements have been completed. This produces opposing cantilevers, as illustrated in Fig. 4a. If the cantilevers are prepared in the fully M1 state by warming from lower temperature to around T_c and are then brought together, the compression produces a domain of R phase in one of them. After retraction, this domain persists only above a certain temperature, and shrinks and disappears below it. We identify this temperature with T_c, the transition temperature at zero stress. By performing the procedure on several devices we obtained the striking result that in every nanobeam T_c was equal to T_{tr}, to within an uncertainty of $\delta T \approx 0.05$ °C governed by temperature fluctuations. We thus conclude that $T_c = T_{tr} = 65.0 \pm 0.1$ °C.

Figure 4b shows the phase diagram of VO_2 inferred from measurements on ten nanobeams (see Supplementary Information for details). In brief, the $P_c(T)|_{ij}$ were deduced from measurements of y_{ij} (i,j = M1, M2, R) versus T as follows. Because the stress in coexistence must take the phase equilibrium value, consideration of the variation of the

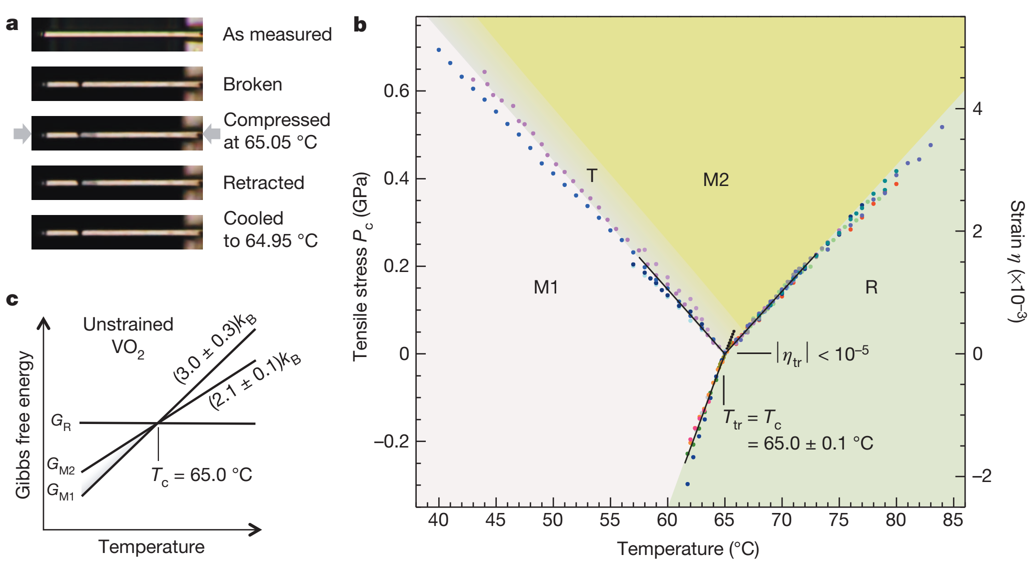


Figure 4 | Phase diagram of VO_2. **a**, The transition temperature T_c at zero stress is measured by finding the temperature above which the metallic phase (darker) becomes stable in a cantilever, as illustrated here (device P8). It is found to be equal to the triple point temperature: $T_c = T_{tr} = 65.0 \pm 0.1$ °C. **b**, Deduced stress–temperature phase diagram. The small black filled circles are for the superheated M1 phase. The grey shaded strip is where a metastable T phase can occur. **c**, The results imply that the free energies of all the phases are degenerate at T_c in unstrained pure VO_2.

strain $\eta = P_c/E$ with T (E is Young's modulus, taken to be 140 GPa for every phase[6]) yields[13]

$$\frac{1}{E}\frac{\partial P_c}{\partial T}\bigg|_{ij} = \frac{d\eta}{dT}\bigg|_{ij} = -\frac{\alpha_{ij}}{L_0}\frac{dy_{ij}}{dT} - \Delta K \quad (1)$$

The first term on the right is the change due to movement of the interface. The second, ΔK, is the thermal expansion mismatch between nanobeam and silicon substrate, which produces a correction of 5–10%. Given that $\eta(T_{tr}) = 0$, equation (1) can be used to derive $\eta(T)$ for each boundary. The deduced phase boundaries are straight, with uncertainties in their slopes of 5–10%, and obey the constraint at the triple point

$$\alpha_{M2M1}\frac{d\eta}{dT}\bigg|_{M2M1} + \alpha_{M1R}\frac{d\eta}{dT}\bigg|_{M1R} = \alpha_{M2R}\frac{d\eta}{dT}\bigg|_{M2R} \quad (2)$$

which is imposed by the Clausius–Clapeyron relations

$$\frac{\partial P_c}{\partial T}\bigg|_{ij} = \frac{S_j - S_i}{b^2(a_i - a_j)} \approx \frac{S_j - S_i}{\alpha_{ij} V} \quad (3)$$

in combination with equation (1). Here S_i is the entropy per vanadium pair in phase i, $b = 4.55$ Å is the base length of the rutile unit cell, and $V = 59$ Å^3 is the rutile unit cell volume. The value of $\partial P_c/\partial T|_{M1R} = 71$ MPa °C^{-1} corresponds to the known latent heat[30] of 1,020 cal per mole formula unit; $\partial P_c/\partial T|_{M2R} = 29$ MPa °C^{-1} corresponds to 710 cal mol^{-1}; and $\partial P_c/\partial T|_{M2M1} = -29$ MPa °C^{-1}. From the results we deduce entropy differences $S_R - S_{M1} = (3.0 \pm 0.3)k_B$ and $S_R - S_{M2} = (2.1 \pm 0.1)k_B$. The equality of T_{tr} and T_c to within $\delta T \approx 0.05$ °C implies that the strain η_{tr} at the triple point is smaller than $\delta T d\eta/dT|_{M2R} = 1.0 \times 10^{-5}$, where $d\eta/dT|_{M2R} = 2.0 \times 10^{-4}$ °C^{-1}, and this is also indicated on the phase diagram. Finally, the finding that the T phase is metastable with respect to M2 is indicated by a grey shaded strip within the M2 stability region.

To stress the implication of these results we sketch in Fig. 4c the T dependence of the Gibbs free energies G_i of the phases of unstrained VO_2, setting $G_R = 0$. The slopes are the entropies $S_i = -dG_i/dT$ at zero stress. Precisely at the MIT the insulating M1 and M2 phases are simultaneously degenerate with the metallic R phase. This and other facts revealed by our measurements are not explained by current models of the transition, but will be crucial ingredients of the correct theory. For example, further development and application of the Landau theory[10] of VO_2 should be prompted by our results. The insights we have gained into this important solid-state phase transition will be critical for both understanding and mastering the MIT in VO_2.

METHODS SUMMARY

VO_2 nanobeams grown by physical vapour transport were transferred onto slots on the micromachined silicon chips by using a micromanipulator and bonded with ultraviolet-curable epoxy (see Supplementary Information). Measurements from ten devices were used, and the temperature was calibrated with the known melting points of gallium and potassium (Supplementary Information). The slot width L (20 or 40 μm) was varied piezoelectrically on a temperature stage under an optical microscope (Supplementary Information).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 April; accepted 24 June 2013.

1. Zylbersztejn, A. & Mott, N. F. Metal–insulator transition in vanadium dioxide. *Phys. Rev. B* **11,** 4383–4395 (1975).
2. Rice, T. M., Launois, H. & Pouget, J. P. Comment on 'VO_2: Peierls or Mott–Hubbard? A view from band theory'. *Phys. Rev. Lett.* **73,** 3042 (1994).
3. Biermann, S., Poteryaev, A., Lichtenstein, A. I. & Georges, A. Dynamical singlets and correlation-assisted Peierls transition in VO_2. *Phys. Rev. Lett.* **94,** 026404 (2005).
4. Eyert, V. VO_2: a novel view from band theory. *Phys. Rev. Lett.* **107,** 016401 (2011).
5. Wu, J. Q. *et al.* Strain-induced self organization of metal–insulator domains in single-crystalline VO_2 nanobeams. *Nano Lett.* **6,** 2313–2317 (2006).
6. Guo, H. *et al.* Mechanics and dynamics of the strain-induced M1–M2 structural phase transition in individual VO_2 nanowires. *Nano Lett.* **11,** 3207–3213 (2011).
7. Atkin, J. M. *et al.* Strain and temperature dependence of the insulating phases of VO_2 near the metal–insulator transition. *Phys. Rev. B* **85,** 020101(R) (2012).
8. Sohn, J. I. *et al.* Surface-stress-induced Mott transition and nature of associated spatial phase transition in single crystalline VO_2 nanowires. *Nano Lett.* **9,** 3392–3397 (2009).
9. Zhang, S. X., Chou, J. Y. & Lauhon, L. J. Direct correlation of structural domain formation with the metal insulator transition in a VO_2 nanobeam. *Nano Lett.* **9,** 4527–4532 (2009).
10. Tselev, A. *et al.* Symmetry relationship and strain-induced transitions between insulating M1 and M2 and metallic R phases of vanadium dioxide. *Nano Lett.* **10,** 4409–4416 (2010).
11. Tselev, A. *et al.* Interplay between ferroelastic and metal–insulator phase transitions in strained quasi-two-dimensional VO_2 nanoplatelets. *Nano Lett.* **10,** 2003–2011 (2010).
12. Cao, J. *et al.* Constant threshold resistivity in the metal–insulator transition of VO_2. *Phys. Rev. B* **82,** 241101(R) (2010).
13. Wei, J., Wang, Z. H., Chen, W. & Cobden, D. H. New aspects of the metal–insulator transition in single-domain vanadium dioxide nanobeams. *Nature Nanotechnol.* **4,** 420–424 (2009).
14. Kasirga, T. S. *et al.* Photoresponse of a strongly correlated material determined by scanning photocurrent microscopy. *Nature Nanotechnol.* **7,** 723–727 (2012).
15. Cao, J. *et al.* Strain engineering and one-dimensional organization of metal–insulator domains in single-crystal vanadium dioxide beams. *Nature Nanotechnol.* **4,** 732–737 (2009).
16. Cao, J. *et al.* Extended mapping and exploration of the vanadium dioxide stress–temperature phase diagram. *Nano Lett.* **10,** 2667–2673 (2010).
17. Podzorov, V., Kim, B. G., Kiryukhin, V., Gershenson, M. E. & Cheong, S. W. Martensitic accommodation strain and the metal–insulator transition in manganites. *Phys. Rev. B* **64,** 140406(R) (2001).
18. Chu, J.-H., Kuo, H.-H., Analytis, J. G. & Fisher, I. R. Divergent nematic susceptibility in an iron arsenide superconductor. *Science* **337,** 710–712 (2012).
19. Becker, M. F., Buckman, A. B. & Walser, R. M. Femtosecond laser excitation of the semiconductor–metal phase transition in VO_2. *Appl. Phys. Lett.* **65,** 1507–1509 (1994).
20. Hilton, D. J. *et al.* Enhanced photosusceptibility near T_c for the light-induced insulator-to-metal phase transition in vanadium dioxide. *Phys. Rev. Lett.* **99,** 226401 (2007).
21. Kubler, C. *et al.* Coherent structural dynamics and electronic correlations during an ultrafast insulator-to-metal phase transition in VO_2. *Phys. Rev. Lett.* **99,** 116401 (2007).
22. Qazilbash, M. M. *et al.* Mott transition in VO_2 revealed by infrared spectroscopy and nano-imaging. *Science* **318,** 1750–1753 (2007).
23. Jones, A. C., Berweger, S., Wei, J., Cobden, D. & Raschke, M. B. Nano-optical investigations of the metal–insulator phase behavior of individual VO_2 microcrystals. *Nano Lett.* **10,** 1574–1581 (2010).
24. Nakano, M. *et al.* Collective bulk carrier delocalization driven by electrostatic surface charge accumulation. *Nature* **487,** 459–462 (2012).
25. Jeong, J. *et al.* Suppression of metal–insulator transition in VO_2 by electric field–induced oxygen vacancy formation. *Science* **339,** 1402–1405 (2013).
26. Wei, J., Ji, H., Guo, W. H., Nevidomskyy, A. H. & Natelson, D. Hydrogen stabilization of metallic vanadium dioxide in single-crystal nanobeams. *Nature Nanotechnol.* **7,** 357–362 (2012).
27. Pouget, J. P., Launois, H., Dhaenens, J. P., Merenda, P. & Rice, T. M. Electron localization induced by uniaxial stress in pure VO_2. *Phys. Rev. Lett.* **35,** 873–875 (1975).
28. Marezio, M., McWhan, B., Dernier, P. D. & Remeika, J. P. Structural aspects of metal–insulator transitions in Cr-doped VO_2. *Phys. Rev. B* **5,** 2541–2551 (1972).
29. Kucharczyk, D. & Niklewski, T. Accurate X-ray determination of the lattice parameters and the thermal expansion coefficients of VO_2 near the transition temperature. *J. Appl. Crystallogr.* **12,** 370–373 (1979).
30. Berglund, C. N. & Guggenheim, H. J. Electronic properties of VO_2 near the semiconductor–metal transition. *Phys. Rev.* **185,** 1022–1033 (1969).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. Spivak, A. Levanyuk and J. Wei for discussions. The silicon chips were patterned at the University of Washington Microfabrication Facility and the Nanofabrication Facility at the University of California, Santa Barbara. This work was supported by the US Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering, award DE-SC0002197.

Author Contributions All authors performed measurements and device fabrication. J.H.P. and J.M.C. designed and constructed the apparatus. D.H.C. conceived and directed the experiments. D.H.C. and J.H.P. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.H.C. (cobden@uw.edu).

doi:10.1038/nature12339

The role of spin in the kinetic control of recombination in organic photovoltaics

Akshay Rao[1], Philip C. Y. Chow[1], Simon Gélinas[1], Cody W. Schlenker[2], Chang-Zhi Li[3], Hin-Lap Yip[3], Alex K.-Y. Jen[2,3], David S. Ginger[2] & Richard H. Friend[1]

In biological complexes, cascade structures promote the spatial separation of photogenerated electrons and holes, preventing their recombination[1]. In contrast, the photogenerated excitons in organic photovoltaic cells are dissociated at a single donor–acceptor heterojunction formed within a de-mixed blend of the donor and acceptor semiconductors[2]. The nanoscale morphology and high charge densities give a high rate of electron–hole encounters, which should in principle result in the formation of spin-triplet excitons, as in organic light-emitting diodes[3]. Although organic photovoltaic cells would have poor quantum efficiencies if every encounter led to recombination, state-of-the-art examples nevertheless demonstrate near-unity quantum efficiency[4]. Here we show that this suppression of recombination arises through the interplay between spin, energetics and delocalization of electronic excitations in organic semiconductors. We use time-resolved spectroscopy to study a series of model high-efficiency polymer–fullerene systems in which the lowest-energy molecular triplet exciton (T_1) for the polymer is lower in energy than the intermolecular charge transfer state. We observe the formation of T_1 states following bimolecular recombination, indicating that encounters of spin-uncorrelated electrons and holes generate charge transfer states with both spin-singlet (^{1}CT) and spin-triplet (^{3}CT) characters. We show that the formation of triplet excitons can be the main loss mechanism in organic photovoltaic cells. But we also find that, even when energetically favoured, the relaxation of ^{3}CT states to T_1 states can be strongly suppressed by wavefunction delocalization, allowing for the dissociation of ^{3}CT states back to free charges, thereby reducing recombination and enhancing device performance. Our results point towards new design rules both for photoconversion systems, enabling the suppression of electron–hole recombination, and for organic light-emitting diodes, avoiding the formation of triplet excitons and enhancing fluorescence efficiency.

The key photophysical processes in an organic photovoltaic cell (OPV) are illustrated in Fig. 1a. In the first step, photogenerated excitons are dissociated by charge transfer across the donor–acceptor interface, leading to either long-range charge separation or the formation of bound interfacial charge transfer states[2] (CTSs). Such bound charge pairs then decay to the ground state by means of geminate recombination. Spin must be taken into account when considering CTSs because they can have either singlet (^{1}CT) or triplet (^{3}CT) spin character, which are almost degenerate in energy[5]. Dissociation of photogenerated singlet excitons leads to the formation of only ^{1}CT states, owing to spin conservation. In contrast, recombination of spin-uncorrelated charges, that is, bimolecular recombination, should lead to the formation of ^{1}CT and ^{3}CT states in a 1:3 ratio, according to spin statistics. Spin-singlet states can recombine to the ground state through either luminescence (which is slow for this intermolecular donor–acceptor process) or non-radiative decay[6]. For ^{3}CT states, decay to the ground state is spin-forbidden and, hence, both radiative and non-radiative processes are very slow. However, if the energy of the T_1 state is less than the ^{3}CT energy (as is required to maximize open circuit voltage, V_{OC}; refs 7, 8), then ^{3}CT can relax to T_1.

The most efficient OPV systems comprise nanoscale (<5 nm) domains of pure fullerene acceptor and domains of fullerene intimately mixed with a polymer donor[9,10]. These length scales are smaller than the Coulomb capture radius, r, in organic semiconductors ($k_B T = e^2/4\pi\varepsilon_0\varepsilon r$, where k_B is Boltzmann's constant, T is the temperature, e is the electron charge, and ε_0 and ε are respectively the vacuum and relative permittivities), which is ~16 nm at room temperature owing to the low dielectric constant of these materials[11] ($\varepsilon \approx 3$–4). This leads to a high rate of electron–hole encounters that could produce Coulombically bound CTSs. This model for recombination and the importance of spin statistics are well established in organic light-emitting diodes, where the formation of (non-luminescent) triplet excitons through bimolecular recombination is a major loss mechanism[3]. Efforts to overcome this problem have focused on the use of metal–organic complexes to induce spin–orbit coupling[12] and, more recently, on the use of low-exchange-energy materials that can promote intersystem crossing from T_1 to S_1 (ref. 13).

In contrast, for OPVs electron–hole encounters have been thought of as terminal recombination events[11,14,15], which, as noted above, is at odds with the high external quantum efficiencies (EQEs) demonstrated in empirically optimized systems[4]. Moreover, the roles of spin and the nature of the intermediate bound CTSs formed after electron–hole capture have not been explored. Here we demonstrate that the recombination of these bound states is mediated not only by energetics, but also by spin and delocalization, allowing for free charges to be reformed from these bound states and thus greatly suppressing recombination.

Figure 1b shows the structures of the two polymers and three fullerene derivatives used as electron donors and, respectively, acceptors in this study. $PC_{60}BM$ (phenyl-C_{60}-butyric acid methyl ester), $PC_{70}BM$ (phenyl-C_{70}-butyric acid methyl ester) and $IC_{60}MA$ (indene–C_{60} monoadduct, referred to as ICMA) are mono-substituted fullerene derivatives[16]. The lowest unoccupied molecular orbital of ICMA is raised by less than 0.1 eV in comparison with $PC_{60}BM$, whereas in $IC_{60}BA$ (indene–C_{60} bisadduct, referred to as ICBA), a bis-substituted derivative, the energy of the lowest unoccupied molecular orbital is raised by about 0.2 eV (refs 17, 18) The donor copolymer PIDT-PhanQ (poly(indacenodithiophene-co-phenanthro[9,10-b]quinoxaline)) was chosen for this study because in it the spectral signatures of charges (hole polarons) and triplets are considerably different[18,19]. As we explain below, this allows us to resolve temporally the interconversion between charges and triplets. It has been recently demonstrated that 1:3 blends of PIDT-PhanQ with $PC_{60}BM$ give excellent photovoltaic performance with internal quantum efficiencies >80% and power conversion efficiencies >4% (ref. 18). In contrast, blends with either ICMA or ICBA give lower performance with power conversion efficiencies of 2.9%. PCPDTBT (poly[2,6-(4,4-bis-(2-ethylhexyl)-4H-cyclopenta[2,1-b;3,

[1]Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK. [2]Department of Chemistry, University of Washington, Seattle, Washington 98195-1700, USA. [3]Department of Materials Science & Engineering, University of Washington, Seattle, Washington 98195-2120, USA.

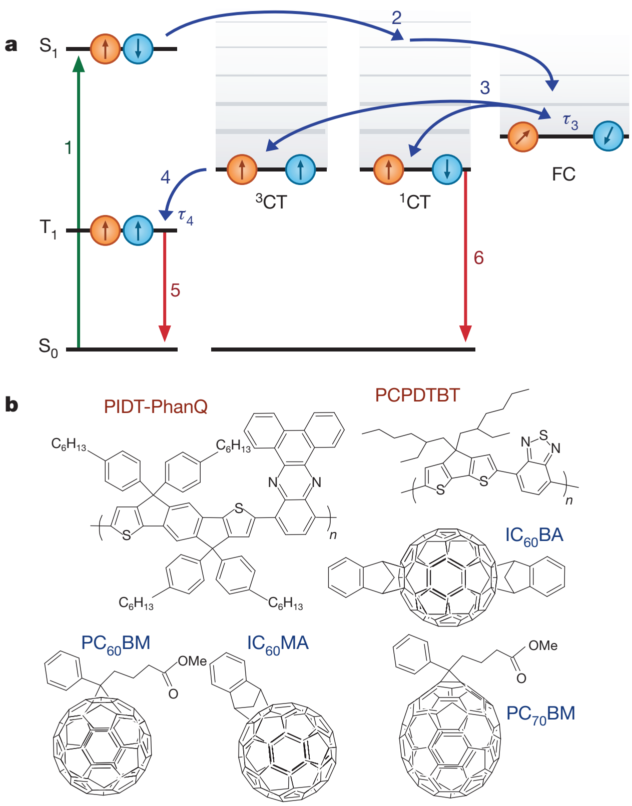


Figure 1 | Photophysical process in an OPV and molecular structures. **a**, State diagram representing the various photophysical processes in an OPV. Conversions between excited state species are shown in blue and recombination channels are shown in red. S_1 and T_1 are the lowest-energy singlet and triplet excitons, respectively. Here we define the CTS energy as the energy of the relaxed, Coulombically bound electron–hole pair across the heterojunction. Process 1: photoexcitation creates a singlet exciton. Process 2: the singlet exciton is ionized at a heterojunction, leading to the formation of ^{1}CT states that separate into free charges (FC) with high efficiency. Process 3: bimolecular recombination of electrons and holes leads to the formation of ^{1}CT and ^{3}CT states in a 1:3 ratio, as mandated by spin statistics. The ^{1}CT state can recombine to the ground state (process 6). Process 4: for the ^{3}CT state, recombination to the ground state is spin-forbidden, but relaxation to the T_1 state is energetically favourable. Process 5: once formed, triplet excitons can return to the ground state through an efficient triplet–charge annihilation channel. Under favourable conditions, as explained in the text, the time required for CTSs to reionize to free charges, τ_3 (process 3), is less than the time required for relaxation to T_1, τ_4 (process 4). Thus, CTSs are recycled back to free charges, leading to a suppression of recombination. **b**, Molecular structures of the donors and acceptors used in this study. Me, methyl.

4-b′]-dithiophene)-alt-4,7-(2,1,3-benzothiadiazole)]) is a widely studied low-bandgap polymer. Despite extensive research, the performance of PCPDTBT:PC_{70}BM blends remains modest, with EQE ≈ 50% (ref. 20). Blends with ICMA or ICBA have even lower performance. Absorption spectra, EQE and current density (J)–voltage (V) curves are provided in Supplementary Information.

For all the studied blends, the energy of the CTS is greater than that of T_1. For PIDT-PhanQ blends, the energies of the CTSs have been previously established using their weak photoluminescence and were found to be 1.31, 1.36 and 1.44 eV for PIDT-PhanQ:PC_{60}BM, PIDT-PhanQ:ICMA and PIDT-PhanQ:ICBA, respectively[18]. The energies of T_1 in PIDT-PhanQ, PCPDTBT and the fullerene derivatives have been established to be 1 (ref. 18), 1 (ref. 21) and 1.5 eV (ref. 22), respectively. The CTS energy of PCPDTBT:PC_{70}BM blends has previously been measured to be 1.2 eV (ref. 23). The CTS energies of PCPDTBT:ICMA and PCPDTBT:ICBA are thus greater than this. Therefore, the molecular triplet exciton of the donor polymer is the lowest-energy excited state for all the studied blends. We note that this is the standard configuration in the current generation of donor–acceptor systems, driven by the need to maximize V_{OC}, which mandates that the charge transfer level lie close to S_1 (refs 7, 8).

Here we investigate thin films of these blends using high-sensitivity transient absorption spectroscopy with a broad spectral and temporal range (Methods). Figure 2a shows the transient absorption spectra of a 1:3 PIDT-PhanQ:PC_{60}BM blend. A broad photoinduced absorption (PIA) feature is formed between the wavelengths 1,100 and 1,500 nm within the instrument response time (2 ns) and decays over several hundred nanoseconds without spectral evolution. The long lifetime of the signal and the fact that it is not observed in pristine films of PIDT-PhanQ (Supplementary Information) rules out PIA by singlet excitons. In contrast, efficient photogeneration of charge is expected in this blend and, thus, the PIA is assigned to charges (hole polarons) on the polymer. Figure 2b shows equivalent spectra for a PIDT-PhanQ:ICMA blend. Here, at the earliest times, the shape of the PIA is similar to that for PIDT-PhanQ:PC_{60}BM (Fig. 2a), but at later times we observe spectral evolution. Between 1,100 and 1,170 nm, the signal decays with time. However, between 1,300 and 1,500 nm the PIA increases for the first 50 ns. The spectrum is also seen to broaden and redshift. This spectral evolution is even more pronounced in the PIDT-PhanQ:ICBA blend shown in Fig. 2c. Thus, unlike the PIDT-PhanQ:PC_{60}BM spectrum, which shows no spectral evolution and is consistent with the decay of a single excited state, the PIDT-PhanQ:ICMA and PIDT-PhanQ:ICBA spectra suggest that a second excited state with a PIA overlapping the PIA of charges is being formed on timescales of tens to hundreds of nanoseconds.

Figure 2e compares the normalized kinetics of PIDT-PhanQ:PC_{60}BM and PIDT-PhanQ:ICBA blends. The PIDT-PhanQ:PC_{60}BM blend (circles) shows no difference between the kinetics of the 1,100–1,200-nm and 1,400–1,500-nm regions, supporting the presence of only one excited-state species. In contrast, for the PIDT-PhanQ:ICBA blend (squares), a large difference in the kinetics of the two regions is observed. The rise time of the low-energy region is much longer than for the higher-energy region, indicating the growth of a second long-lived excited state species on nanosecond timescales.

Figure 2f compares the normalized kinetics of the 1,400–1,500-nm region in PIDT-PhanQ:ICBA for different values of pulse fluence. A clear dependence on pulse fluence is observed, with rise times (to the signal maximum) as large as 80 ns. Similar fluence dependence for the rise time is not observed for the 1,100–1,200-nm region (Supplementary Information), with the signal maximum occurring within the rise time of the instrument. The rise time of the 1,400–1,500-nm region is also fluence dependent in PIDT-PhanQ:ICMA but not in PIDT-PhanQ:PC_{60}BM (Supplementary Information). This fluence dependence in ICBA and ICMA blends indicates that the second excited-state species growing in is formed by bimolecular processes.

The overlapping spectra of the excited states make the analysis of their kinetics difficult. To overcome this problem, we use a genetic algorithm[24] that allows us to extract the individual spectra and kinetics from the data set (Methods). Figure 2d shows the two spectra (solid lines) that the algorithm extracts from the PIDT-PhanQ:ICBA spectrum in Fig. 2c. The spectrum in blue is the charge (hole polaron) and the one in red is the triplet exciton on PIDT-PhanQ. These assignments are based on previous continuous-wave PIA experiments[18] as well as early-time transient absorption measurements (Supplementary Information).

From the spectra and kinetics presented in Fig. 2, we can now observe that charges are formed within the instrument response time (2 ns) in all blends. For the 1:3 PIDT-PhanQ:PCB_{60}M blend presented in Fig. 2a, charges then decay on a 1-μs timescale and no triplet formation is observed. For PIDT-PhanQ:ICMA and PIDT-PhanQ:ICBA, triplet excitons are formed through bimolecular recombination on nanosecond timescales before decaying.

Figure 3a, b shows the kinetics, at various fluences, extracted from the genetic-algorithm-based global analysis for PIDT-PhanQ:ICMA

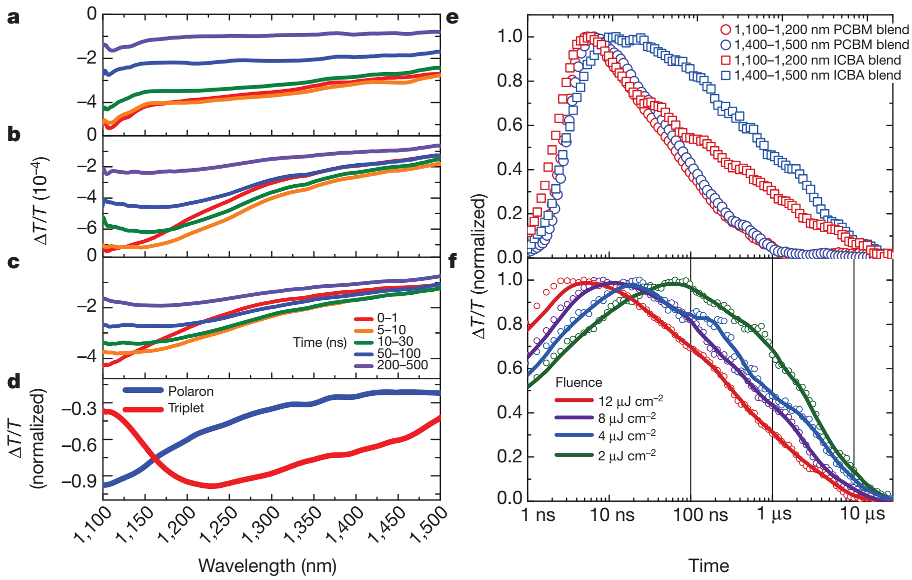


Figure 2 | Excited-state spectra and kinetics for PIDT-PhanQ blends. **a–c**, Temporal evolution of the transient absorption spectra for 1:3 blend ratios of PIDT-PhanQ:PC_{60}BM (**a**), PIDT-PhanQ:ICMA (**b**) and PIDT-PhanQ:ICBA (**c**). Samples were excited with an excitation fluence of 8 μJ cm^{-2}. Temporal slices are averaged over the indicated time periods and smoothed. The PC_{60}BM blend (**a**) shows only a slow decay with no spectral evolution. In contrast, the ICMA (**b**) and ICBA (**c**) blends show evolution up to 100 ns, indicating the growth of a new excited state species. **d**, The spectra extracted from the genetic algorithm analysis of the ICBA blend (**c**) showing the triplet and charge polaron spectra. **e**, Comparison of kinetics of PIDT-PhanQ:PC_{60}BM (circles) and PIDT-PhanQ:ICBA (squares). Both the high-energy (red) and low-energy (blue) regions for PIDT-PhanQ:PC_{60}BM decay on the same timescale, indicating the presence of only one excited-state species. In contrast, for PIDT-PhanQ:ICBA the regions have divergent kinetics, indicating the presence of multiple excited-state species. **f**, Fluence dependence of the low-energy region (1,400–1,500 nm) for PIDT-PhanQ:ICBA. The fluence-dependent growth of the feature demonstrates that the second excited-state species, triplets, are formed through bimolecular processes.

and PIDT-PhanQ:ICBA. The extracted kinetics clearly demonstrate that triplets grow in as charges decay. We consider that the primary decay channel for triplets is triplet–charge annihilation, owing to the high charge densities present, and model the time evolution of the system (Fig. 3a, b, solid lines) using the equation

$$\frac{dN_T}{dt} = -\alpha\frac{dp}{dt} - \beta N_T p \qquad (1)$$

where p is the charge concentration, N_T is the triplet concentration, α is the fraction of decaying charges that form triplets and β is the rate constant for triplet–charge annihilation.

Values of β vary by a factor of two with fluence, and at a fluence of 2 μJ cm^{-2} for the PIDT-PhanQ:ICBA blend we obtain a value of 0.58 for α and a value of 2.2×10^{-10} cm^3 s^{-1} for β (Supplementary Information). This demonstrates that a large fraction of charges undergo bimolecular recombination, mediated by ^{3}CT, to form triplet excitons. Once formed, triplets are quickly quenched as a result of triplet–charge annihilation, as indicated by the high value of β. This is important: given sufficient time, triplets could be re-ionized through thermal excitation to CTSs. However, the presence of a strong triplet–charge annihilation channel means that recombination to triplets is a terminal loss and makes it a major loss pathway in OPVs.

We now turn to the question of whether the time taken for relaxation from ^{3}CT to T_1 (process 4 in Fig. 1a, with an associated timescale τ_4) is fast and, if not, whether there are competing processes for the decay of ^{3}CT. As noted earlier, for all PIDT-PhanQ:fullerene blends the charge transfer energy is greater than the T_1 energy, making relaxation from ^{3}CT to T_1 energetically favoured. However, for the more

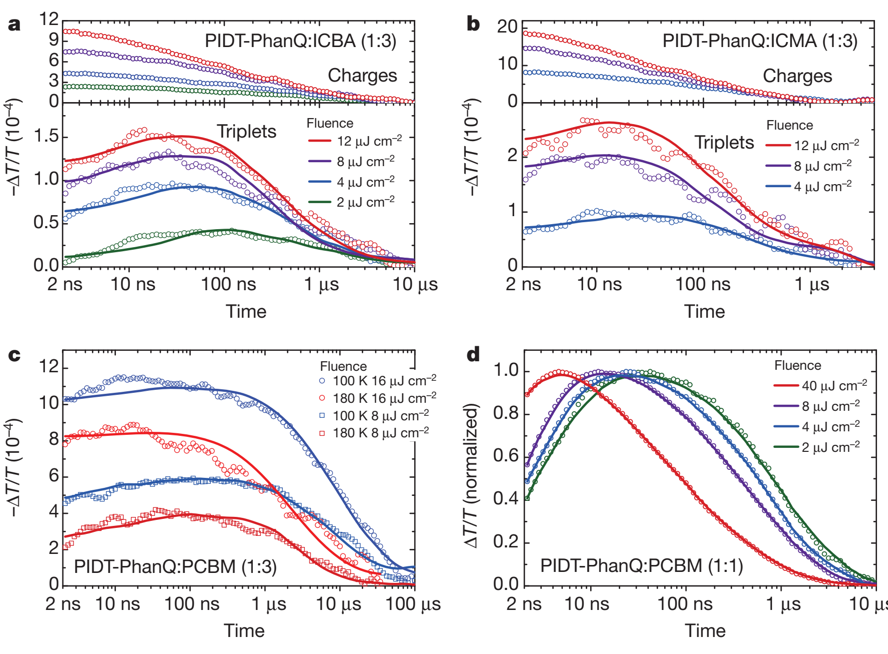


Figure 3 | Triplet and charge kinetics for PIDT-PhanQ blends. **a**, **b**, Charge and triplet dynamics for PIDT-PhanQ:ICBA (**a**) and PIDT-PhanQ:ICMA (**b**), extracted from the genetic algorithm analysis. Charges are formed within the instrument response time in all cases. The growth of triplets is fluence dependent, with a maximum population attained at later times for lower fluences. The solid lines are fits of the experimental data using the model described in the text. **c**, Temperature-dependent triplet dynamics (extracted using the genetic algorithm analysis) for a 1:3 PIDT-PhanQ:PC_{60}BM sample. Open symbols show the experimental data and solid lines are fits using the model described in the text. The maximum triplet population is formed at later times at lower fluences and lower temperatures, consistent with a bimolecular-diffusion-dependent process. **d**, Room-temperature (297 K) fluence-dependent triplet dynamics (extracted using the genetic algorithm analysis) for a 1:1 PIDT-PhanQ:PC_{60}BM blend spun from chloroform. In contrast to the 1:3 PIDT-PhanQ:PC_{60}BM blend (Fig. 2a), this blend shows room-temperature triplet formation, which, as indicated by the fluence dependence, stems from bimolecular processes. Solid lines are guides to the eye.

efficient 1:3 PIDT-PhanQ:$PC_{60}BM$ blend, no triplet formation is observed at room temperature (Fig. 2a). But at low temperatures (<240 K), bimolecular triplet formation is observed in this blend (Fig. 3c; temperature-dependent kinetics of the raw data are provided in Supplementary Information). The solid lines are fits using the model described in equation (1). This result suggests that there is a thermally activated process that competes with relaxation to T_1. We consider this process to be the dissociation of ^3CT back to free charges. This is based on the fact that no other excited-state species are observed for this system (Fig. 2). Thus, at high temperatures (>240 K) the dissociation of ^3CT back to free charges (process 3 in Fig. 1a, with an associated timescale τ_3) out-competes relaxation of ^3CT to T_1; that is, $\tau_4 > \tau_3$. At lower temperatures, this dissociation process is suppressed, such that $\tau_4 < \tau_3$, leading to a build-up of triplet excitons (Fig. 3c).

The above result raises the question of why triplet formation is observed in ICBA and ICMA blends at room temperature but is out-competed by dissociation back to free charges in the 1:3 PIDT-PhanQ:$PC_{60}BM$ blend. As noted above, the charge transfer levels of the ICMA and $PC_{60}BM$ blends are within 50 meV of each other and, hence, a simple energetics argument is unlikely to explain this difference. Our previous work on CTSs formed at early times through the ionization of excitons at heterojunctions suggested that their dissociation was mediated by charge wavefunction delocalization[2]. We propose that the same mechanism is applicable to CTSs formed through bimolecular recombination. It is known that PCBM forms large aggregates efficiently, in contrast to other fullerenes, and that aggregation aids charge separation[25]. This effect is most probably due to delocalization of the electron wavefunction over the PCBM aggregates—fullerenes forming smaller aggregates would lead to more localized electron wavefunctions. This would imply that CTSs formed through recombination were more loosely bound in PIDT-PhanQ:$PC_{60}BM$ (1:3 blends) than in PIDT-PhanQ:ICMA and PIDT-PhanQ:ICBA and were thus more susceptible to dissociation back to free charges. To test this hypothesis, we study the recombination dynamics in a 1:1 PIDT-PhanQ:$PC_{60}BM$ blend spun from chloroform. The lower fullerene concentration and low-boiling-point solvent lead to a more intimate blend and arrest the growth of large fullerene aggregates. This is confirmed by grazing-incidence small-angle X-ray scattering measurements (Supplementary Information), which also show formation of smaller aggregates in the ICMA and ICBA blends than in the 1:3 PIDT-PhanQ:$PC_{60}BM$ blend. Bimolecular triplet formation is observed in the 1:1 PIDT-PhanQ:$PC_{60}BM$ blend (Fig. 3d), which shows the normalized fluence dependence of the triplets (raw data and charge dynamics are shown in Supplementary Information). Thus, by disrupting fullerene aggregation and, hence, charge delocalization, we make $\tau_4 < \tau_3$. This result confirms that delocalization has a crucial role in recombination.

To generalize the above results, we now study PCPDTBT blends. Figure 4a shows the evolution of the transient absorption spectrum of a 1:2 PCPDTBT:ICBA blend. A broad PIA feature between 1,175 and 1,550 nm is formed within the instrument response time (2 ns), and its peak blueshifts from 1,300 to 1,275 nm over tens of nanoseconds. A similar blueshift was observed in films of PCPDTBT:$PC_{70}BM$ and PCPDTBT:ICMA (Supplementary Information). The triplet spectrum extracted from a genetic algorithm analysis of the blends (Fig. 4b, solid red line) shows excellent agreement with the measured triplet spectrum in neat PCPDTBT (Fig. 4b, dashed red line; see Methods). The triplet peak at higher energy, with respect to the charge, explains the blueshifting of the transient absorption spectrum in Fig. 4a as triplets grow in.

Figure 4c shows the fluence dependence of the charge and triplet for a PCPDTBT:$PC_{70}BM$ film, similar to those shown in Fig. 3a, b. The solid lines are fits to the experimental data obtained using equation (1), and support the general applicability of the presented model. The result also explains why PCPDTBT blends have only modest EQEs[20], with recombination to triplets being a major loss mechanism even for the $PC_{70}BM$ blend.

On the basis of these results, we can now propose a new photophysical model of recombination in OPVs, summarized in Fig. 1a. Within working devices, the high charge densities present[26] (10^{16}–10^{17} cm^{-3}) lead to bimolecular electron–hole capture events forming CTSs with both spin-singlet and spin-triplet character, ^1CT and ^3CT. Recombination from ^3CT back to the ground state is spin-forbidden, but in most systems T_1 is lower in energy than CT, such that energetic relaxation to bound triplet excitons is favourable (with an associated time τ_4). The high charge densities then result in rapid quenching of these triplets. Thus, for most blends two recombination channels exist, one by means of triplets and the other through the radiative and non-radiative relaxation of the ^1CT states. However, as demonstrated in the PIDT-PhanQ:$PC_{60}BM$ system, when the acceptor is well ordered (encouraging wavefunction delocalization) the reionization of CTSs back to free charges (with an associated time τ_3) can occur faster than relaxation to T_1 ($\tau_4 > \tau_3$). In this case, the triplet recombination pathway is turned off, leaving only recombination through the singlet channel.

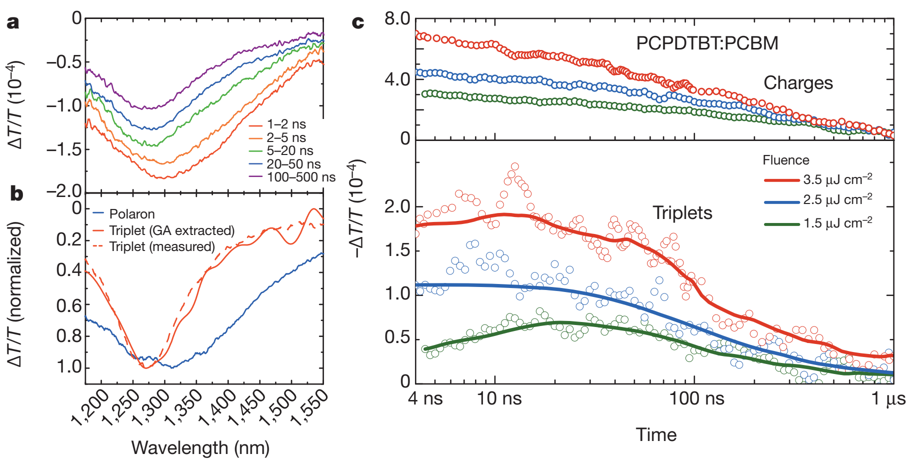


Figure 4 | Triplet and charge kinetics for PCPDTBT blends. a, Temporal evolution of the transient absorption spectrum for PCPDTBT:ICBA, excited with an excitation fluence of 2 μJ cm^{-2}. Temporal slices are averaged over the indicated time periods and smoothed. A blueshift of the spectra from a peak at 1,300 nm to 1,275 nm can be seen over the first 100 ns. **b**, Triplet spectrum extracted from the genetic algorithm (GA) analysis (solid red line) and that measured by doping a PCPDTBT thin film with a triplet sensitizer (dashed red line). The blue line shows the charge spectrum as measured 50 ps after photoexcitation, which is sufficient time for charge generation but not enough for triplet formation to begin. **c**, Charge and triplet dynamics (circles) for PCPDTBT:$PC_{70}BM$, extracted from the global genetic algorithm analysis, analogous to those shown in Fig. 3a, b. Charges are formed within the instrument response time in all cases. The growth of triplets is fluence dependent, with a maximum population attained at later times for lower fluences. The solid lines are fits of the experimental data using the model described in the text.

For working OPV devices, bimolecular recombination controls the shape of the J–V curve[26], with extraction of charges at the electrodes in competition with recombination. Recombination to triplets can proceed faster than extraction, and we observe triplet formation in working devices even under short-circuit conditions (Supplementary Information). As the voltage increases from short-circuit conditions towards V_{OC}, charge densities and extraction times increase, leading to higher bimolecular recombination losses. The film measurements here represent the case of V_{OC}, in which there is no extraction and recombination of all charges. We also note that any bimolecular recombination process that is non-radiative must reduce efficiency below the Shockley–Queisser limit[27], so avoidance of triplet formation is always desirable.

We note finally that the recombination current in OPVs is analogous to the injection current in organic light-emitting diodes, where electrons and holes with uncorrelated spins are injected from the electrode and recombine within the active layer. Recent efforts to minimize losses due to the formation of non-radiative triplets have focused on manipulating energetic levels such that triplet states are higher in energy than CTSs[27], or on finding systems with very low exchange energies[13]. However, these approaches can impose restrictive design criteria on materials and rely on inherently slow intersystem crossing from triplet to singlet. What we show here is that the introduction of weakly bound CTSs makes it possible to shut off recombination to non-radiative triplets, even when they are the lowest-energy excited state, and to achieve efficient recombination through the singlet channel. This insight opens a new route to high-efficiency fluorescent organic light-emitting diodes.

METHODS SUMMARY

PCPDTBT, ICMA and ICBA were obtained from 1-material, and $PC_{60}BM$ and $PC_{70}BM$ from Nano-C. PIDT-PhanQ was synthesized as described previously[18,19].

For the PIDT-PhanQ:fullerene thin-film samples, 1:3 polymer:fullerene blends ($20\,mg\,ml^{-1}$ in dichlorobenzene) were spun on fused-silica substrates. The PIDT-PhanQ:$PC_{60}BM$ thin film discussed in Fig. 3d was spun from a 1:1 blend ($20\,mg\,ml^{-1}$ in chloroform). For the PCPDTBT:fullerene thin-film samples, 1:2 polymer:fullerene blends ($30\,mg\,ml^{-1}$ in chlorobenzene) were spun on fused-silica substrates.

For transient absorption measurements, 90-fs pulses generated in a Ti:sapphire amplifier system (Spectra-Physics Solstice) operating at 1 kHz were used. The broadband probe beam was generated in a home-built non-collinear optical parametric amplifier. Pump pulses were generated using a frequency-doubled, q-switched Nd:YVO_4 laser (532 nm). Delay times from 1 ns to 100 μs were achieved by synchronizing the pump laser with the probe pulse using an electronic delay generator. Samples were measured in a dynamic vacuum ($<1 \times 10^{-5}$ mbar).

Full Methods and any associated references are available in the online version of the paper.

Received 4 February; accepted 29 May 2013.
Published online 7 August 2013.

1. Blankenship, R. E. *Molecular Mechanisms of Photosynthesis* ch. 7 (Blackwell Science, 2001).
2. Bakulin, A. A. *et al.* The role of driving energy and delocalized states for charge separation in organic semiconductors. *Science* **335,** 1340–1344 (2012).
3. Wallikewitz, B. H., Kabra, D., Gelinas, S. & Friend, R. H. Triplet dynamics in fluorescent polymer light-emitting diodes. *Phys. Rev. B* **85,** 045209 (2012).
4. Park, S. H. *et al.* Bulk heterojunction solar cells with internal quantum efficiency approaching 100%. *Nature Photon.* **3,** 297–302 (2009).
5. Hu, B., Yan, L. & Shao, M. Magnetic-field effects in organic semiconducting materials and devices. *Adv. Mater.* **21,** 1500–1516 (2009).
6. Morteani, A. C., Sreearunothai, P., Herz, L. M., Friend, R. H. & Silva, C. Exciton regeneration at polymeric semiconductor heterojunctions. *Phys. Rev. Lett.* **92,** 247402 (2004).
7. Veldman, D., Meskers, S. C. J. & Janssen, R. A. J. The energy of charge-transfer states in electron donor–acceptor blends: insight into the energy losses in organic solar cells. *Adv. Funct. Mater.* **19,** 1939–1948 (2009).
8. Vandewal, K., Tvingstedt, K., Gadisa, A., Inganas, O. & Manca, J. V. On the origin of the open-circuit voltage of polymer-fullerene solar cells. *Nature Mater.* **8,** 904–909 (2009).
9. Hammond, M. R. *et al.* Molecular order in high-efficiency polymer/fullerene bulk heterojunction solar cells. *ACS Nano* **5,** 8248–8257 (2011).
10. Lou, S. J. *et al.* Effects of additives on the morphology of solution phase aggregates formed by active layer components of high-efficiency organic solar cells. *J. Am. Chem. Soc.* **133,** 20661–20663 (2011).
11. Hilczer, M. & Tachiya, M. Unified theory of geminate and bulk electron-hole recombination in organic solar cells. *J. Phys. Chem. C* **114,** 6808–6813 (2010).
12. Baldo, M. A. *et al.* Highly efficient phosphorescent emission from organic electroluminescent devices. *Nature* **395,** 151–154 (1998).
13. Uoyama, H., Goushi, K., Shizu, K., Nomura, H. & Adachi, C. Highly efficient organic light-emitting diodes from delayed fluorescence. *Nature* **492,** 234–238 (2012).
14. Deibel, C., Baumann, A. & Dyakonov, V. Polaron recombination in pristine and annealed bulk heterojunction solar cells. *Appl. Phys. Lett.* **93,** 163303 (2008).
15. Koster, L. J. A., Mihailetchi, V. D. & Blom, P. W. M. Bimolecular recombination in polymer/fullerene bulk heterojunction solar cells. *Appl. Phys. Lett.* **88,** 052104 (2006).
16. Guangjin, Z., Youjun, H. & Yongfang, L. 6.5% efficiency of polymer solar cells based on poly(3-hexylthiophene) and indene-C_{60} bisadduct by device optimization. *Adv. Mater.* **22,** 4355–4358 (2010).
17. Hoke, E. T. *et al.* The role of electron affinity in determining whether fullerenes catalyze or inhibit photooxidation of polymers for solar cells. *Adv. Energy Mater.* **2,** 1351–1357 (2012).
18. Schlenker, C. W. *et al.* Polymer triplet energy levels need not limit photocurrent collection in organic solar cells. *J. Am. Chem. Soc.* **134,** 19661–19668 (2012).
19. Zhang, Y. *et al.* Indacenodithiophene and quinoxaline-based conjugated polymers for highly efficient polymer solar cells. *Chem. Mater.* **23,** 2289–2291 (2011).
20. Peet, J. *et al.* Efficiency enhancement in low-bandgap polymer solar cells by processing with alkane dithiols. *Nature Mater.* **6,** 497–500 (2007).
21. Di Nuzzo, D. *et al.* Improved film morphology reduces charge carrier recombination into the triplet excited state in a small bandgap polymer-fullerene photovoltaic cell. *Adv. Mater.* **22,** 4321–4324 (2010).
22. Soon, Y. W. *et al.* Energy versus electron transfer in organic solar cells: a comparison of the photophysics of two indenofluorene:fullerene blend films. *Chem. Sci.* **2,** 1111–1120 (2011).
23. Scharber, M. C. *et al.* Charge transfer excitons in low band gap polymer based solar cells and the role of processing additives. *Energy Environ. Sci.* **4,** 5077–5083 (2011).
24. Gélinas, S. *et al.* The binding energy of charge-transfer excitons localized at polymeric semiconductor heterojunctions. *J. Phys. Chem. C* **115,** 7114–7119 (2011).
25. Jamieson, F. C. *et al.* Fullerene crystallisation as a key driver of charge separation in polymer/fullerene bulk heterojunction solar cells. *Chem. Sci.* **3,** 485–492 (2012).
26. Credgington, D., Hamilton, R., Atienzar, P., Nelson, J. & Durrant, J. R. Non-geminate recombination as the primary determinant of open-circuit voltage in polythiophene:fullerene blend solar cells: an analysis of the influence of device processing conditions. *Adv. Funct. Mater.* **21,** 2744–2753 (2011).
27. Shockley, W. & Queisser, H. J. Detailed balance limit of efficiency of p-n junction solar cells. *J. Appl. Phys.* **32,** 510–519 (1961).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank N. Greenham for discussions. A.R. thanks Corpus Christi College, Cambridge for a Research Fellowship. S.G. thanks Fonds Québécois de Recherche sur la Nature et les Technologies for funding. This work is supported by the EPSRC and the Winton Programme for the Physics of Sustainability. C.W.S. was supported by the National Science Foundation (DMR-1215753). D.S.G., C.-Z.L., H.-L.Y. and A.K.-Y.J. acknowledge support from the Office of Naval Research (N00014-11-1-0300). Some of the work was done at the UW NanoTech User Facility, a member of the NSF National Nanotechnology Infrastructure Network. We thank J. Richards and D. Pozzo for performing grazing-incidence small-angle X-ray scattering measurements, S. Williams for transmission electron microscopy and G. Shao for help with atomic force microscopy measurements.

Author Contributions A.R. and P.C.Y.C. performed the time-resolved measurements. S.G. developed the numerical methods. A.R., P.C.Y.C. and S.G. analysed the data. C.W.S. and D.S.G. had the idea for the structural and steady-state spectroscopic measurements. C.-Z.L. synthesized PIDT-PhanQ. H.-L.Y. and A.K.-Y.J. had the idea for the molecular design of PIDT-PhanQ. R.H.F. supervised the work. A.R., P.C.Y.C., S.G. and R.H.F. wrote the manuscript. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.H.F. (rhf10@cam.ac.uk).

METHODS

Transient absorption spectroscopy. In this technique, a pump pulse generates photoexcitations within the film, which are then studied at some later time using a broadband probe pulse. Although transient absorption has been widely used to study the photophysics of OPV blends, previous measurements have been severally limited by three factors: first, insufficient temporal range, typically a maximum of 2-ns delay between pump and probe; second, limited spectral range and a lack of broadband probes, which hinders the observation of dynamic interactions between excitations; and, last, insufficient sensitivity, which mandates the use of high excitation densities to create large signals. Here we overcome these problems by using broad temporal (up to 1 ms) and spectral windows (up to 1,500 nm) and high sensitivity (better than 5×10^{-6}). Although broad temporal[25] and spectral[28] windows have previously been achieved, all three requirements have not been met simultaneously before.

The temporal window is created by the use of an electrically delayed pump pulse and allows for the study of long-lived charges and triplet excitons. This was achieved by synchronizing the pump laser (a frequency-doubled, q-switched Nd:YVO_4 laser (532 nm) with 800-ps pulse width; AOT-YVO-25QSPX, Advanced Optical Technologies) with the probe pulse using an electronic delay generator (SRS DG535, Stanford Research Systems).

In conjugated polymers, local geometrical relaxation around charges (polaron formation) causes rearrangement of energy levels, bringing states into the semiconductor gap and giving rise to strong optical transitions in the range 700–1,500 nm (ref. 2). The absorption bands of singlet and triplet excitons are also found to lie in the near infrared, making a broadband spectral window necessary to track the evolution of the excited-state species. To generate these probe pulses, a portion of the output of a Ti:sapphire amplifier system (Spectra-Physics Solstice) operating at 1 kHz was used to pump a home-built non-collinear optical parametric amplifier modelled after ref. 29. The probe beam was split and a portion passed through a region of the sample not affected by the pump, so that laser fluctuations could be normalized. The probe and reference signals were dispersed in a spectrometer (Andor, Shamrock SR-303i) and detected using a pair of 16-bit, 512-pixel linear image sensors (Hamamatsu).

For short time measurements (pump–probe delay, <2 ns), the excitation pulses were generated by a TOPAS optical parametric amplifier (Light Conversion; 300-fs pulse width) seeded with a portion of the amplifier output, and the pump was delayed using a mechanical delay stage (Newport). Every second pump pulse was omitted electronically when using the q-switched source or, for short-time measurements, using a mechanical chopper. Data acquisition at 1 kHz was enabled by a custom-built board from Entwicklungsbüro Stresing. The differential transmission ($\Delta T/T$) was calculated after accumulating and averaging 1,000 pump-on and pump-off shots for each data point.

High stability of the probe beam, the use of a reference pulse to correct for shot-to-shot variation and the ability to read out every shot allows for a high signal-to-noise ratio. The high sensitivity of the experiment is essential as it allows us to probe the dynamics of systems when the excitation densities are similar to solar illumination conditions[26] (10^{16}–10^{17} excitations cm^{-3}; see Supplementary Information for calculation of excitation densities). At higher excitation densities, bimolecular exciton–exciton and exciton–charge annihilation processes can dominate, creating artefacts and making such measurements unreliable indicators of device operation under AM1.5G illumination[30].

All measurements were carried out under a dynamic vacuum ($<1 \times 10^{-5}$ mbar). Data obtained was then smoothed using a moving average filter in MATLAB. The step size of the filter was small to avoid losing spectral and temporal accuracy. For measurement of the triplet spectra of pristine PCPDTBT, the film was doped with an iridium complex that enhances the intersystem crossing rate, leading to a high triplet population. The measured triplet spectrum was found to be in good agreement with a previous measurement[28].

Numerical methods. To deconvolve the overlapping signatures of individual excited states and obtain their kinetics, we use numerical methods based on a genetic algorithm. The full details of this approach can be found elsewhere[24]. In summary, the algorithm starts by generating a large population of random spectra, which it then breeds into successive generations of offspring using a 'survival of the fittest' approach. Once the fitness stops improving with newer generations, the best spectra are returned as the optimized solution. For a given solution, the fitness is calculated as the inverse of the sum of squared residual, with a penalty added for non-physical results. The parents are selected using a tournament method with adaptive crossover and the offspring generated using a Gaussian-function mask (of random parameters) as the relative weight of each parent spectrum at a given wavelength.

28. Etzold, F. *et al.* The effect of solvent additives on morphology and excited-state dynamics in PCPDTBT:PCBM photovoltaic blends. *J. Am. Chem. Soc.* **134,** 10569–10583 (2012).
29. Cirmi, G. *et al.* Few-optical-cycle pulses in the near-infrared from a noncollinear optical parametric amplifier. *Opt. Lett.* **32,** 2396–2398 (2007).
30. Hodgkiss, J. M. *et al.* Exciton-charge annihilation in organic semiconductor films. *Adv. Funct. Mater.* **22,** 1567–1577 (2012).

LETTER

doi:10.1038/nature12376

Onset of deglacial warming in West Antarctica driven by local orbital forcing

WAIS Divide Project Members*

The cause of warming in the Southern Hemisphere during the most recent deglaciation remains a matter of debate[1,2]. Hypotheses for a Northern Hemisphere trigger, through oceanic redistributions of heat, are based in part on the abrupt onset of warming seen in East Antarctic ice cores and dated to 18,000 years ago, which is several thousand years after high-latitude Northern Hemisphere summer insolation intensity began increasing from its minimum, approximately 24,000 years ago[3,4]. An alternative explanation is that local solar insolation changes cause the Southern Hemisphere to warm independently[2,5]. Here we present results from a new, annually resolved ice-core record from West Antarctica that reconciles these two views. The records show that 18,000 years ago snow accumulation in West Antarctica began increasing, coincident with increasing carbon dioxide concentrations, warming in East Antarctica and cooling in the Northern Hemisphere[6] associated with an abrupt decrease in Atlantic meridional overturning circulation[7]. However, significant warming in West Antarctica began at least 2,000 years earlier. Circum-Antarctic sea-ice decline, driven by increasing local insolation, is the likely cause of this warming. The marine-influenced West Antarctic records suggest a more active role for the Southern Ocean in the onset of deglaciation than is inferred from ice cores in the East Antarctic interior, which are largely isolated from sea-ice changes.

Exceptional records of Southern Hemisphere climate change come from Antarctic ice cores[2,6,7]. Most of these records are from high-altitude sites on the East Antarctic plateau. Questions about the reliability of the two previous deep West Antarctic ice-core records result in those records often being excluded from reconstructions of Antarctic climate[4,8]. Because the climate of West Antarctica is distinct from that of interior East Antarctica, the exclusion of West Antarctic records may result in an incomplete picture of past Antarctic and Southern Ocean climate change. Interior West Antarctica is lower in elevation and more subject to the influence of marine air masses than interior East Antarctica, which is surrounded by a steep topographic slope[9,10]. Marine-influenced locations are important because they more directly reflect atmospheric conditions resulting from changes in ocean circulation and sea ice. However, ice-core records from coastal sites are often difficult to interpret because of complicated ice-flow and elevation histories. The West Antarctic Ice Sheet (WAIS) Divide ice core (WDC), in central West Antarctica, is unique in coming from a location that has experienced minimal elevation change[11], is strongly influenced by marine conditions[9] and has a relatively high snow-accumulation rate, making it possible to obtain an accurately dated record with high temporal resolution.

Drilling of WDC was completed in December 2011 to a depth of 3,405 m. Drilling was halted ~50 m above the bedrock to avoid contaminating the basal water system. WDC is situated 24 km west of the Ross–Amundsen ice-flow divide and 160 km east of the Byrd ice-core site (Supplementary Fig. 1). The elevation is 1,766 m; the present-day snow accumulation rate is 22 cm yr^{-1} (ice equivalent) and the average temperature is approximately $-30\,^{\circ}$C. The age of the oldest recovered ice is ~68 kyr. The WDC06A-7 timescale is based on the identification of annual layers to 29.6 kyr ago using primarily electrical measurements (Methods). To validate WDC06A-7, we compare times of abrupt changes in atmospheric methane concentration (Supplementary Information) with the Greenland Ice Core Chronology 2005[12] (GICC05). We also compare the methane variations in WDC with abrupt changes in a speleothem δ^{18}O record from Hulu Cave, China. The difference in age between the ice and gas at a given depth is calculated using a steady-state firn-densification model and is always less than 500 yr. The age differences between WDC06A-7 and GICC05 and between WDC06A-7 and the Hulu Cave timescale are much less than the independent timescale uncertainties (Supplementary Fig. 6).

We interpret δ^{18}O of ice (Methods) as annual-mean surface air temperature, as supported by independent estimates of temperature from borehole thermometry[13]. WDC has many similarities with other records (Fig. 1) and resolves Antarctic Isotope Maximum (AIM) events clearly. The late Holocene WDC record shows cooling, suggesting that the increase in δ^{18}O at Byrd over the past few thousand years resulted from ice advection and thinning[11]. The abrupt increase in δ^{18}O ~22 kyr ago at Siple Dome is not observed at WDC. The AIM1 peak and the subsequent Antarctic Cold Reversal (ACR; 14.5–12.9 kyr ago) are more pronounced in WDC than at Byrd and Siple Dome, possibly owing to discontinuous sampling of the Byrd core and thinning of Siple Dome.

The most rapid warming at WDC occurred after the ACR and culminated at AIM0. The timing of AIM0 is difficult to define because it is composed of two peaks, one 11.95 kyr ago and the other 11.6 kyr ago. The ice accumulation rate at WDC increased abruptly by 37% in the 400 yr between 12.0 and 11.6 kyr ago (Supplementary Fig. 2). The increase in ice accumulation with little change in δ^{18}O shows that the accumulation rate is not controlled strictly by temperature. Abrupt changes in accumulation cannot be recognized in most other Antarctic ice cores because their timescales lack sufficient resolution; it is thus unknown whether this event is specific to WDC or whether accumulation increased abruptly over a larger portion of Antarctica.

The coldest period at WDC was between 28 and 22 kyr ago and was interrupted by AIM2, a 1,000-yr warm period between 24 and 23 kyr ago. AIM2 is also prominent in the EPICA Dronning Maud Land (EDML) ice core[7] but is muted or nearly absent in other East Antarctic records[14] (Fig. 1). Other West Antarctic cores also record AIM2, although the low resolution of the Byrd core and the abrupt δ^{18}O increase 22 kyr ago in the Siple Dome core have made this feature difficult to discern. AIM2 illustrates the spatial heterogeneity of Antarctic climate variability during the coldest part of the glacial period.

To investigate deglacial warming across the Antarctic continent, we use a sliding Wilcoxon rank-sum test (Fig. 2) to identify times of significant change in the δ^{18}O records of WDC, EDML and the EPICA Dome C ice core[6] (EDC); we convert the EDC δD record to δ^{18}O using $\delta^{18}\mathrm{O} = (\delta\mathrm{D} - 10)/8$. The WDC and EDC timescales can be aligned at a ~150-yr-long acid deposition event[15,16], which eliminates the relative age uncertainty at 18 kyr ago. The rank sum test reveals three important features: gradual deglacial warming at WDC was punctuated by periods of more rapid change; the most abrupt warming

*Lists of participants and their affiliations appear at the end of the paper.

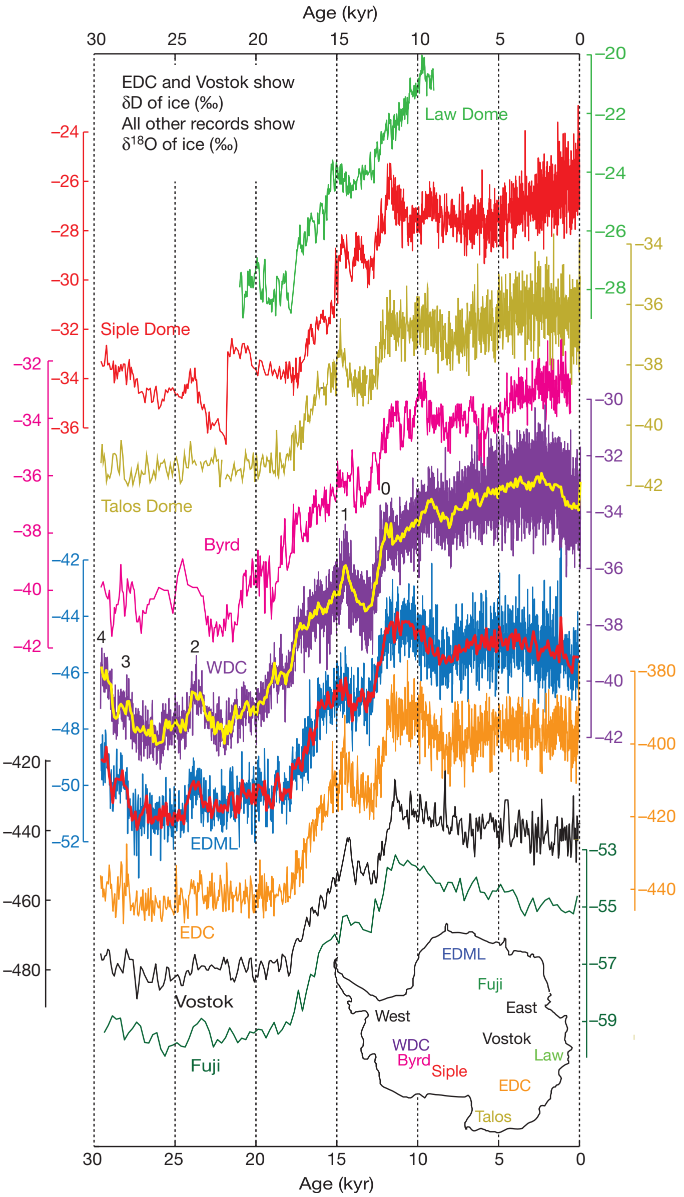


Figure 1 | Antarctic Isotope Records. Water isotope ratios from nine Antarctic ice cores. Inset, outline of Antarctica with the ice-core locations: Law Dome[27] (light green), Siple Dome[28] (red), Byrd[29] (pink), Talos Dome[14] (khaki), WDC (purple, WDC06A-7 timescale), EDML[7] (blue), EDC[6] (orange), Dome Fuji[2] (dark green), Vostok[30] (black). Taylor Dome is not plotted because of timescale uncertainties. All records are at original resolution. Thick lines for WDC and EDML are 50-yr averages. EDML, EDC and Vostok use the Lemieux-Dudon[31] timescale. Numbers above the WDC curve indicate AIM events. ($\delta^{18}O = (^{18}O/^{16}O)_{sample}/(^{18}O/^{16}O)_{VSMOW} - 1$ and $\delta D = (^{2}H/^{1}H)_{sample}/(^{2}H/^{1}H)_{VSMOW} - 1$).)

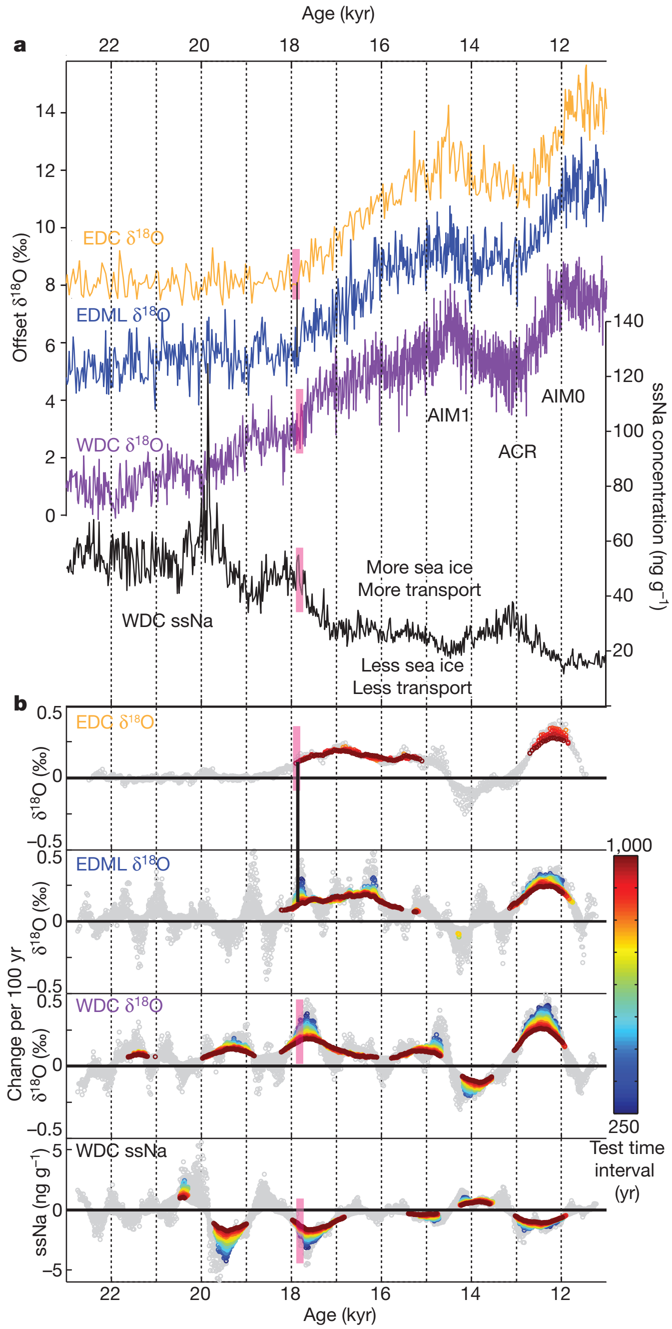


Figure 2 | Timing of rapid change in Antarctica. a, Water isotope ratios ($\delta^{18}O$, purple) and ssNa concentrations (black) from WDC on WDC06A-7. EDML[7] (blue) and EDC[6] (orange) $\delta^{18}O$ use the Lemieux-Dudon[31] timescale. Constants have been subtracted from $\delta^{18}O$ records for plotting. Magenta boxes indicate a 150-yr acid deposition event; the black line between EDC and EDML is a volcanic tie point (Methods). ssNa is plotted as 25-yr median values. **b**, Rate of change for $\delta^{18}O$ at WDC, EDML, and EDC and ssNa at WDC. A Wilcoxon rank-sum test (Methods) is used to determine significance. Significant rates of change are coloured by test time interval; rates of change that are not significant are coloured grey.

began at the 18-kyr-ago acid deposition event; and significant warming at WDC began by 20 kyr ago, at least 2,000 yr before significant warming at EDML and EDC.

Further insight into deglacial warming at WDC is gained by investigating the sea-salt sodium (ssNa) record (Methods). Debate remains about whether ssNa on millennial timescales reflects primarily sea-ice production or the strength of atmospheric circulation[17]. In the Amundsen and Ross seas, changes in sea ice and atmospheric circulation are coupled because atmospheric forcing is the dominant control on sea-ice concentration[18]. We interpret ssNa as a proxy for sea-ice extent and a marker of marine changes (Supplementary Information). The rank-sum test reveals that each rapid increase in $\delta^{18}O$, indicating warming, was accompanied by a decrease in ssNa, suggesting less sea ice (Fig. 2). Consistent with this, the decrease in $\delta^{18}O$ during the ACR was accompanied by an increase in ssNa.

The accumulation rate at WDC was inferred without assuming a relationship with $\delta^{18}O$ or temperature (Methods). Although uncertainty in the annual-layer interpretation and ice-flow history used to determine the accumulation rate precludes a statistical assessment comparable to that used for the $\delta^{18}O$ and ssNa records, results suggest that an initial increase in accumulation occurred between 18.5 and

17 kyr ago (Fig. 3), consistent with the rapid warming 18 kyr ago. This also coincides with evidence for changes in Southern Ocean upwelling[19], atmospheric carbon dioxide concentration[8,20] and Atlantic meridional overturning circulation[5] (AMOC). The accumulation increase probably results from more frequent or stronger moisture-bearing storms penetrating into West Antarctica. This supports a southward shift[21] or intensification[22] of the mid-latitude westerly storm track, and is consistent with the hypothesis of a decrease in AMOC leading to Southern Hemisphere warming and Northern Hemisphere cooling[3]—the 'bipolar seesaw'.

Both the WDC and the lower-resolution Byrd ice-core records show that warming in West Antarctica began before the decrease in AMOC that has been invoked to explain Southern Hemisphere warming[4,19]. The most significant early warming at WDC occurred between 20 and 18.8 kyr ago, although a period of significant warming also occurred between 22 and 21.5 kyr ago. The magnitude of the warming at WDC before 18 kyr ago is much greater than at EDML or EDC; linear regression of $\delta^{18}O$ between 22 and 18 kyr ago shows that it increased by 2.2‰ at WDC, by 0.4‰ at EDML and by 0.1‰ at EDC (Fig. 2). It is very unlikely that the 2.2‰ increase at WDC can be attributed to elevation change; this magnitude of isotope change would require more than 200 m of ice-sheet thinning, twice the amount of thinning that occurred during the Holocene epoch when the grounding line retreated hundreds of kilometres (Supplementary Information). The subdued warming at EDML and the lack of warming at EDC are consistent with the lack of clear AIM2 signals in some East Antarctic cores, and suggest that cores from the East Antarctic plateau do not capture the full magnitude of Southern Hemisphere climate variability.

There is evidence that warming at WDC before 18 kyr ago is due to decreasing sea ice. The ssNa at WDC began to decrease 20 kyr ago, probably as a result of both decreasing sea-ice extent and decreasing strength of transport from changes in atmospheric circulation. A marine record from the southwest Atlantic Ocean indicates that significant summer and winter sea-ice retreat began before 22 kyr ago[23]. Furthermore, a reduction in sea-ice extent can explain the different magnitude of warming among ice-core sites before 18 kyr ago. The high East Antarctic plateau is largely isolated from coastal changes because the local marine air masses do not have the energy to rise above the steep coastal escarpment[10].

To illustrate the variable sensitivity of different areas in Antarctica to changes in sea-ice extent, we used an atmospheric general circulation model[24]. Using Last Glacial Maximum (LGM) sea surface temperature and sea-ice boundary conditions from a fully coupled model run[25], we performed a control run of the ECHAM4.6 atmospheric model with the LGM sea-ice extent and a comparison run with reduced sea-ice extent (Supplementary Information). Sea surface temperatures are prescribed; the atmospheric circulation therefore responds to the change in sea-ice extent but the sea-ice extent is not further affected by the changes in atmospheric circulation. The magnitude of sea-ice retreat is consistent with evidence for reduced sea ice in the southwest Atlantic between 22 and 18 kyr ago[23]. In response to the sea-ice retreat, all of West Antarctica and coastal East Antarctica is enriched in precipitation-weighted $\delta^{18}O$, whereas interior East Antarctica is little changed or is depleted (Fig. 4). The positive $\delta^{18}O$ anomalies probably extend unrealistically far into the East Antarctic interior because of the low-resolution topography in the climate model. Although the details of the spatial pattern of $\delta^{18}O$ anomalies are dependent on model resolution and on the specified boundary conditions, the greater

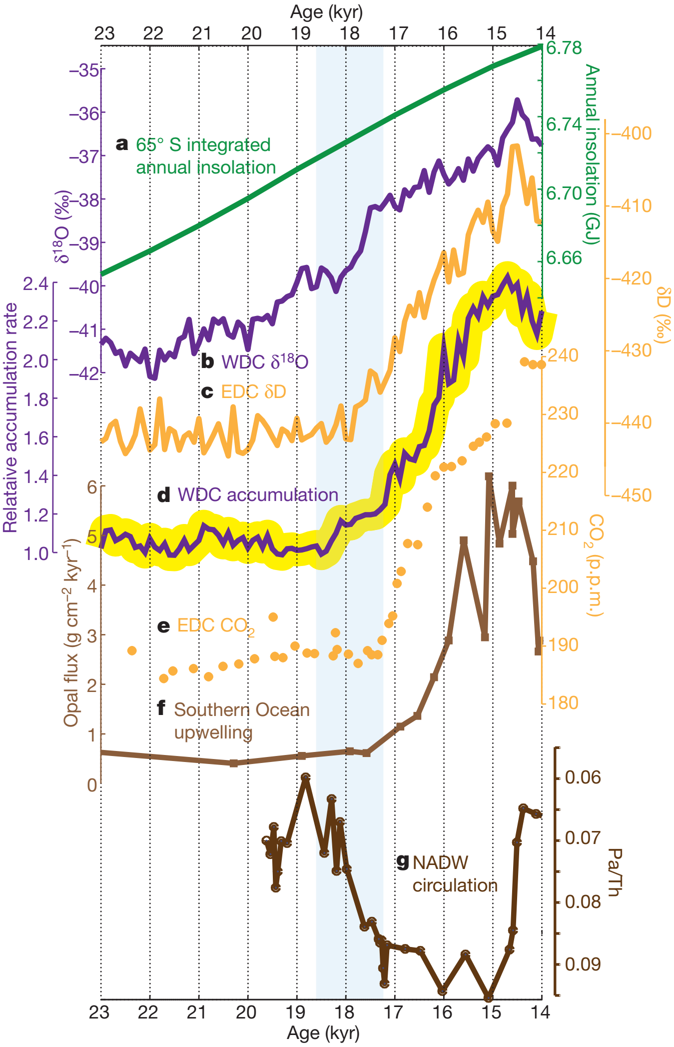


Figure 3 | Global records of deglaciation. **a**, Integrated annual insolation at latitude 65° S. **b**, 100-yr averages of $\delta^{18}O$ at WDC on WDC06A-7. **c**, 100-yr averages of δD at EDC[6] on the Lemieux-Dudon[31] timescale. **d**, Relative accumulation rate (normalized to the mean value between 19.5 and 18.5 kyr ago) at WDC. Yellow shading is the uncertainty in identifying annual layers (Methods). **e**, Atmospheric CO_2 concentration[20] from EDC on the Lemieux-Dudon[31] timescale. **f**, Opal flux[19], a proxy for upwelling, from ocean sediment core TNO57-13-4PC in the South Atlantic. **g**, Pa/Th, a proxy for North Atlantic Deep Water (NADW) circulation[5], from sediment core GCC5. Blue shading indicates a period with relatively abrupt changes in all palaeoclimate records (**b**–**g**).

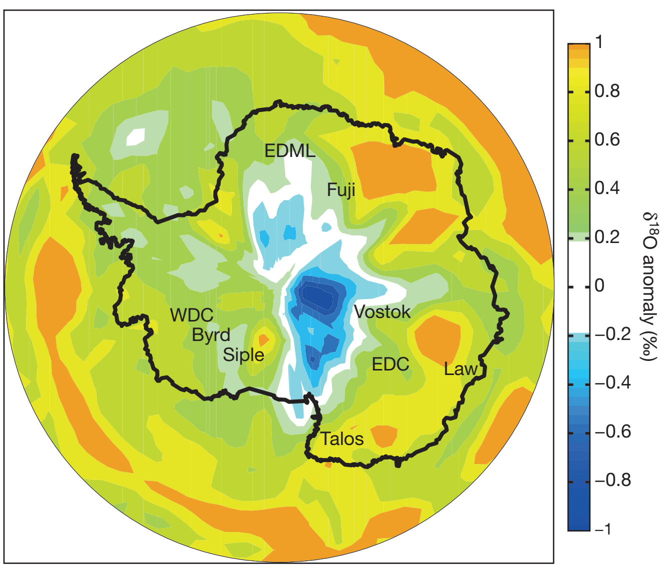


Figure 4 | Antarctic $\delta^{18}O$ response to sea-ice decrease. Response of precipitation-weighted $\delta^{18}O$ to an approximately zonally symmetric southward displacement of the sea-ice edge (Supplementary Fig. 9) in the ECHAM4.6 climate model run with LGM boundary conditions.

sensitivity of the WAIS Divide region to sea-ice decline compared with locations in interior East Antarctica is clear.

Local orbital forcing is a likely cause of the inferred sea-ice change. Integrated annual insolation at latitude 65° S increased by 1% between 22 and 18 kyr ago. The additional annual insolation is 60 MJ m^{-2}, which is enough to melt 5 cm m^{-2} of sea ice assuming an albedo of 0.75. The increase in integrated summer insolation, where summer is defined as days with insolation above a threshold[26] of 275 W m^{-2}, is greater than the total annual increase (Supplementary Fig. 10). Thus, the increase comes in summer, when it is most likely to be absorbed by low-albedo open water. The summer duration also begins increasing at 23 kyr ago; longer summers and shorter winters may also contribute to the decrease in sea-ice extent[1]. The effect of an increase in insolation would be amplified by the sea-ice/albedo feedback.

The abrupt onset of East Antarctic warming[4,8], increasing CO_2 (ref. 20) and decreasing AMOC[5] 18 kyr ago has supported the view that deglaciation in the Southern Hemisphere is primarily a response to changes in the Northern Hemisphere[3]. Yet the evidence of warming in West Antarctica and corresponding evidence for sea-ice decline in the southeast Atlantic[23] show that climate changes were ongoing in the Southern Ocean before 18 kyr ago, supporting an important role for local orbital forcing[1]. Warming in the high latitudes of both hemispheres before 18 kyr ago implies little change in the interhemispheric temperature gradient that largely determines the position of the intertropical convergence zone and the position and intensity of the mid-latitude westerlies[21,22]. We propose that when Northern Hemisphere cooling occurred ~18 kyr ago, coupled with an already-warming Southern Hemisphere, the intertropical convergence zone and mid-latitude westerlies shifted southwards in response. The increased wind stress in the Southern Ocean drove upwelling, venting of CO_2 from the deep ocean[19] and warming in both West Antarctica and East Antarctica. The new WDC record thus reveals an active role for the Southern Hemisphere in initiating global deglaciation.

METHODS SUMMARY

The WDC06A-7 timescale is based on measurements of sulphur, sodium, black carbon and electrical conductivity above 577 m (to 2,358 yr before AD 1950), and primarily on electrical measurements below 577 m. Using atmospheric methane as a stratigraphic marker, WDC06A-7 and GICC05 agree to within 100 ± 200 yr at the three abrupt changes between 14.7 and 11.7 kyr ago; WDC06A-7 is older by 500 ± 600 yr at 24 kyr ago, by 250 ± 300 yr at 28 kyr ago and by 350 ± 250 yr at 29 kyr ago (Supplementary Fig. 6). WDC06A-7 agrees within the uncertainties with the Hulu Cave timescale and is older by 50 ± 300 yr at 28 kyr ago and by 100 ± 300 yr at 29 kyr ago.

We measured δ^{18}O at a resolution of 0.5 m using laser spectroscopy with calibration to Vienna Standard Mean Ocean Water (VSMOW). We report ssNa concentration rather than flux because wet deposition dominates at higher accumulation rates. The accumulation-rate record was derived independently from the stable-isotope record using a one-dimensional ice-flow model to calculate the thinning function.

Periods of significant change in δ^{18}O and ssNa are identified with a sliding, non-parametric Wilcoxon rank-sum test. The data were averaged to 25-yr resolution for WDC and EDML, and to 50-yr resolution for EDC. We tested pairs of adjacent blocks of data against the null hypothesis of equal medians, performing the test at all points along the record. We assessed change on multiple timescales using a range of block sizes corresponding to time intervals of 250–1,000 yr for WDC and EDML and 500–1,000 yr for EDC. We used an effective 95% a-posteriori confidence requirement; the critical significance level (p) was determined as $1 - 0.95^{1/N}$ where N is the number of test realizations.

We used the ECHAM4.6 atmospheric general circulation model at T42 resolution (2.8° by 2.8°) with 19 vertical levels and glacial sea surface temperature boundary conditions.

Full Methods and any associated references are available in the online version of the paper.

Received 4 January; accepted 12 June 2013.
Published online 14 August 2013.

1. Huybers, P. & Denton, G. Antarctic temperature at orbital timescales controlled by local summer duration. *Nature Geosci.* **1,** 787–792 (2008).
2. Kawamura, K. *et al.* Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. *Nature* **448,** 912–916 (2007).
3. Clark, P. U., Pisias, N. G., Stocker, T. F. & Weaver, A. J. The role of the thermohaline circulation in abrupt climate change. *Nature* **415,** 863–869 (2002).
4. Shakun, J. D. *et al.* Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. *Nature* **484,** 49–54 (2012).
5. McManus, J. F., Francois, R., Gherardi, J. M., Keigwin, L. D. & Brown-Leger, S. Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* **428,** 834–837 (2004).
6. EPICA community members. Eight glacial cycles from an Antarctic ice core. *Nature* **429,** 623–628 (2004).
7. EPICA. Community Members. One-to-one coupling of glacial climate variability in Greenland and Antarctica. *Nature* **444,** 195–198 (2006).
8. Parrenin, F. *et al.* Synchronous change of atmospheric CO_2 and Antarctic temperature during the last deglacial warming. *Science* **339,** 1060–1063 (2013).
9. Nicolas, J. P. & Bromwich, D. H. Climate of West Antarctica and influence of marine air intrusions. *J. Clim.* **24,** 49–67 (2011).
10. Noone, D. & Simmonds, I. Sea ice control of water isotope transport to Antarctica and implications for ice core interpretation. *J. Geophys. Res.* **109,** D07105 (2004).
11. Steig, E. J. *et al.* in *The West Antarctic Ice Sheet: Behavior and Environment* Vol. 77 (eds Alley, R. & Bindschadler, R.) 75–90 (American Geophysical Union, 2001).
12. Svensson, A. *et al.* A 60,000 year Greenland stratigraphic ice core chronology. *Clim. Past* **4,** 47–57 (2008).
13. Steig, E. J. *et al.* Recent climate and ice-sheet changes in West Antarctica compared with the past 2,000 years. *Nature Geosci.* **6,** 372–375 (2013).
14. Stenni, B. *et al.* Expression of the bipolar see-saw in Antarctic climate records during the last deglaciation. *Nature Geosci.* **4,** 46–49 (2011).
15. Hammer, C. U., Clausen, H. B. & Langway, C. C. 50,000 years of recorded global volcanism. *Clim. Change* **35,** 1–15 (1997).
16. Schwander, J. *et al.* A tentative chronology for the EPICA Dome Concordia ice core. *Geophys. Res. Lett.* **28,** 4243–4246 (2001).
17. Wolff, E. W., Rankin, A. M. & Rothlisberger, R. An ice core indicator of Antarctic sea ice production? *Geophys. Res. Lett.* **30,** 2158 (2003).
18. Holland, P. R. & Kwok, R. Wind-driven trends in Antarctic sea-ice drift. *Nature Geosci.* **5,** 872–875 (2012).
19. Anderson, R. F. *et al.* Wind-driven upwelling in the Southern Ocean and the deglacial rise in atmospheric CO_2. *Science* **323,** 1443–1448 (2009).
20. Monnin, E. *et al.* Atmospheric CO_2 concentrations over the last glacial termination. *Science* **291,** 112–114 (2001).
21. Toggweiler, J. R., Russell, J. L. & Carson, S. R. Midlatitude westerlies, atmospheric CO2, and climate change during the ice ages. *Paleoceanography* **21,** PA2005 (2006).
22. Lee, S. Y., Chiang, J. C. H., Matsumoto, K. & Tokos, K. S. Southern Ocean wind response to North Atlantic cooling and the rise in atmospheric CO_2: modeling perspective and paleoceanographic implications. *Paleoceanography* **26,** PA1214 (2011).
23. Collins, L. G., Pike, J., Allen, C. S. & Hodgson, D. A. High-resolution reconstruction of southwest Atlantic sea-ice and its role in the carbon cycle during marine isotope stages 3 and 2. *Paleoceanography* **27,** PA3217 (2012).
24. Roeckner, E. *et al. The Atmospheric General Circulation Model ECHAM-4: Model Description and Simulation of Present-Day Climate*. Report No. 218 90 (Max-Planck-Institut für Meteorologie, 1996).
25. Braconnot, P. *et al.* Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum - Part 1: experiments and large-scale features. *Clim. Past* **3,** 261–277 (2007).
26. Huybers, P. Early Pleistocene glacial cycles and the integrated summer insolation forcing. *Science* **313,** 508–511 (2006).
27. Pedro, J. B. *et al.* The last deglaciation: timing the bipolar seesaw. *Clim. Past* **7,** 671–683 (2011).
28. Brook, E. J. *et al.* Timing of millennial-scale climate change at Siple Dome, West Antarctica, during the last glacial period. *Quat. Sci. Rev.* **24,** 1333–1343 (2005).
29. Blunier, T. & Brook, E. J. Timing of millennial-scale climate change in Antarctica and Greenland during the last glacial period. *Science* **291,** 109–112 (2001).
30. Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399,** 429–436 (1999).
31. Lemieux-Dudon, B. *et al.* Consistent dating for Antarctic and Greenland ice cores. *Quat. Sci. Rev.* **29,** 8–20 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by US National Science Foundation (NSF). The authors appreciate the support of the WAIS Divide Science Coordination Office (M. Twickler and J. Souney) for the collection and distribution of the WAIS Divide ice core; Ice Drilling and Design and Operations (K. Dahnert) for drilling; the National Ice Core Laboratory (B. Bencivengo) for curating the core; Raytheon Polar Services (M. Kippenhan) for logistics support in Antarctica; and the 109th New York Air National Guard for airlift in Antarctica. We also thank C. Buizert and S. Marcott for discussions. The following individual NSF grants supported this work: 0944197 (E.D.W., H. Conway); 1043092, 0537930 (E.J.S.); 0944348, 0944191, 0440817, 0440819, 0230396 (K.C.T.); 0538427, 0839093 (J.R.M.); 1043518 (E.J.B.); 1043500 (T.S.); 05379853, 1043167 (J.W.C.W.); 1043528, 0539578 (R.B.A.); 0539232 (K.M.C., G.D.C.); 1103403 (R.L.E., H. Conway); 0739780 (R.E.); 0637211 (G.H.); 0538553, 0839066 (J.C.-D.), 0538657, 1043421 (J.P.S.); 1043313 (M.K.S.); 0801490 (G.J.W). Other support came from a NASA NESSF award (T.J.F.), the USGS Climate and Land Use Change Program (G.D.C., J.J.F.), the National Natural Science Foundation of China (41230524 to H. Cheng) and the Singapore National Research Foundation (NRFF2011-08 to X.W.).

Author Contributions The manuscript was written by T.J.F., E.J.S. and B.R.M. K.C.T. organized the WAIS Divide Project. T.J.F., K.C.T and T.J.P. made the electrical measurements and developed the electrical timescale with K.C.M. E.J.S., J.W.C.W., A.J.S., P.N., B.H.V. and S.W.S. measured the stable-isotope record. J.R.M., M.S., O.J.M. and R.E. developed the chemistry timescale and measured Na. E.J.B., T.S., L.E.M., J.S.E. and J.E.L. made the methane measurements. G.D.C. and K.M.C. measured the borehole temperature profile. J.C.-D. and D.F. provided an independent timescale for the brittle ice. Q.D., S.W.S. and E.J.S. performed the climate modelling. T.J.F., E.D.W., H. Conway and K.M.C. performed the ice-flow modelling to determine the accumulation rate. H. Cheng, R.L.E., X.W., J.P.S. and T.J.F. made comparisons with the Hulu cave timescale. M.K.S., J.J.F., J.M.F., D.E.V. and R.B.A. examined the physical properties of the core. W.M., J.J. and N.M. designed the drill. G.H. designed core-processing techniques. A.J.O., B.H.V., D.E.V., K.C.T., T.J.P. and G.J.W. led collection and processing of the core in the field.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.J.F. (tjfudge@uw.edu).

WAIS Divide Project Members T. J. Fudge[1], Eric J. Steig[1,2], Bradley R. Markle[1], Spruce W. Schoenemann[1], Qinghua Ding[1,2], Kendrick C. Taylor[3], Joseph R. McConnell[3], Edward J. Brook[4], Todd Sowers[5], James W. C. White[6,7], Richard B. Alley[5,8], Hai Cheng[9,10], Gary D. Clow[11], Jihong Cole-Dai[12], Howard Conway[1], Kurt M. Cuffey[13], Jon S. Edwards[4], R. Lawrence Edwards[10], Ross Edwards[14], John M. Fegyveresi[5,8], David Ferris[12], Joan J. Fitzpatrick[15], Jay Johnson[16], Geoffrey Hargreaves[17], James E. Lee[4], Olivia J. Maselli[3], William Mason[18], Kenneth C. McGwire[3], Logan E. Mitchell[4], Nicolai Mortensen[16], Peter Neff[1,19], Anais J. Orsi[20], Trevor J. Popp[21], Andrew J. Schauer[1], Jeffrey P. Severinghaus[20], Michael Sigl[3], Matthew K. Spencer[22], Bruce H. Vaughn[7], Donald E. Voigt[5,8], Edwin D. Waddington[1], Xianfeng Wang[23] & Gifford J. Wong[24]

Affiliations for participants: [1]Department of Earth and Space Sciences, University of Washington, Seattle, Washington 98195, USA. [2]Quaternary Research Center, University of Washington, Seattle, Washington 98195, USA. [3]Desert Research Institute, Nevada System of Higher Education, Reno, Nevada 89512, USA. [4]College of Earth, Ocean and Atmospheric Sciences Oregon State University, Corvallis, Oregon 97331, USA. [5]Earth and Environmental Systems Institute, Pennsylvania State University, University Park, Pennsylvania 16802, USA. [6]Department of Geological Sciences and Department of Environmental Studies, Boulder, Colorado 80309, USA. [7]INSTAAR, University of Colorado, Boulder, Colorado 80309, USA. [8]Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA. [9]Institute of Global Environmental Change, Xi'an Jiaotong University, Xi'an 710049, China. [10]Department of Earth Sciences, University of Minnesota, Minneapolis, Minnesota 55455, USA. [11]US Geological Survey, Geosciences and Environmental Change Science Center, Lakewood, Colorado 80225, USA. [12]Department of Chemistry and Biochemistry, South Dakota State University, Brookings, South Dakota 57007, USA. [13]Department of Geography, University of California-Berkeley, Berkeley 94720, USA. [14]Department of Imaging and Applied Physics, Curtin University, Perth, Western Australia 6102, Australia. [15]US Geological Survey, Denver, Colorado 80225, USA. [16]Ice Drilling Design and Operations, Space Science Engineering Center, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. [17]US Geologic Survey, National Ice Core Laboratory, Denver, Colorado 80225, USA. [18]EMECH Designs, Brooklyn, Wisconsin 53521, USA. [19]Antarctic Research Centre, Victoria University of Wellington, Wellington 6012, New Zealand. [20]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92037, USA. [21]Centre for Ice and Climate, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen, Denmark. [22]Department of Geology and Physics, Lake Superior State University, Sault Ste Marie, Michigan 49783, USA. [23]Earth Observatory of Singapore, Nanyang Technological University, Singapore 639798. [24]Department of Earth Sciences, Dartmouth College, Hanover, New Hampshire 03755, USA.

METHODS

Stable-isotope measurements of ice. Water isotope analyses were by laser spectroscopy[32] at the University of Washington. Values of $\delta^{18}O$ represent the deviation from Vienna Standard Mean Ocean Water (VSMOW) normalized[11] to the VSMOW-SLAP standards and reported in per mil (‰). The precision of the measurements is better than 0.1‰. The data have not been corrected for advection, elevation, or mean seawater $\delta^{18}O$.

Accumulation rates. The accumulation-rate record was derived independently from the stable-isotope record using an ice-flow model to calculate the thinning function. We use a transient one-dimensional ice-flow model to compute the vertical-velocity profile:

$$w(z) = -\left(\dot{b} - \dot{m} - \dot{H}\right)\psi(z) - \dot{m} - \left(\frac{\rho_i}{\rho(z)} - 1\right)\dot{b} \tag{1}$$

Here z is the height above the bed, $\dot{b}$ is the accumulation rate, $\dot{m}$ is the melt rate, $\dot{H}$ is the rate of ice-thickness change, ρ_i is the density of ice, $\rho(z)$ is the density profile and $\psi(z)$ is the vertical velocity shape function computed as

$$\psi(z) = \frac{(f_B z + (1/2)(1-f_B)(z^2/h))}{(H-(h/2)(1-f_B))} \quad \text{for } h \geq z > 0$$

$$\psi(z) = \frac{(z-(h/2)(1-f_B))}{(H-(h/2)(1-f_B))} \quad \text{for } H \geq z > h$$

following ref. 33. Here h is the distance above bedrock of the Dansgaard–Johnsen[34] kink height, f_B is the fraction of the horizontal surface velocity due to sliding over the bed and H is the ice thickness. Firn compaction is incorporated through the rightmost term in equation (1) and assumes a density profile that does not vary with time.

A constant ice thickness was specified because the thickness change near the divide was probably small (~100 m) and the timing of thickening and thinning is not well constrained; a 100-m thickness change would alter the inferred accumulation rate by ~3%. A constant basal melt rate of 1 cm yr^{-1} and non-divide flow conditions, represented by a Dansgaard–Johnsen kink height of 0.2H, were assumed. We also prescribed a sliding fraction of 0.5 of the surface velocity, which approximates effects of both basal sliding and enhanced shear near the bed, neither of which is well constrained. To assess the possible range of inferred accumulation rates, we also used sliding fractions of 0.15 and 0.9 (Supplementary Fig. 2). The inferred accumulation rate was only slightly affected for the Holocene part of the record but differed by up to 16% for the oldest part of the record (29.6 kyr ago). Because the thinning function varies smoothly, the uncertainty in the timing of the changes in accumulation rate is only weakly affected by the uncertainty in the magnitude of the accumulation rate. The main uncertainty in identifying the timing of accumulation rate changes is the uncertainty in the timescale itself. During the deglacial transition, the uncertainty in the interpretation is estimated at 8%. The yellow shading in Fig. 3 shows this uncertainty.

WDC06A-7 timescale. The WDC06A-7 timescale is based on high-resolution (<1 cm) measurements of sulphur, sodium, black carbon and electrical conductivity (ECM) above 577 m (2,358 yr before present (BP; AD 1950); ref. 35). Below 577 m, WDC06A-7 is based primarily on electrical measurements: di-electrical profiling was used for the brittle ice from 577 to 1,300 m (to 6,063 yr BP). Alternating-current ECM measurements were used from 1,300 to 1,955 m (to 11,589 yr BP) and both alternating-current and direct-current ECM measurements were used below 1,955 m. The interpretation was stopped at 2,800 m because the expression of annual layers becomes less consistent, suggesting that all years may not be easily recognized.

The upper 577 m of the timescale has been compared with volcanic horizons dated on multiple other timescales[35]; the uncertainty at 2,358 yr BP is ±19 yr. For the remainder of the timescale, we assigned an uncertainty based on a qualitative assessment of the clarity of the annual layers. For ice from 577 to 2,020 m (2–12 kyr ago), we estimated a 2% uncertainty based on comparisons between the ECM and chemical (Na, SO_4) interpretations between 577 and 1,300 m, which agreed to within 1% (Supplementary Fig. 4). The estimated uncertainty increased during the deglacial transition owing to both thinner layers and a less pronounced seasonal cycle. We compared the annual-layer interpretation of the ECM records in an 800-yr overlap section (1,940–2,020-m depth, corresponding to 11.4–12.2 kyr ago) with various high-resolution chemistry records (sodium and sulphur). We found overall good agreement (19 yr more in the ECM-only interpretation) but did observe a tendency for the ECM record to 'split' one annual peak into two small peaks. We used this knowledge in the annual-layer interpretation of the ECM record. We increased the uncertainty to 4% between 2,020 and 2,300 m (12.2–15.5 kyr ago) and to 8% between 2,300 and 2,500 m (15.5–20 kyr ago). The glacial period had a stronger annual-layer signal than the transition, and we estimate a 6% uncertainty for the rest of the glacial. The 150-yr acid deposition event, first identified in the Byrd ice core[15], was found in WDC at depths of 2,421.75 to 2,427.25 m. Because there is consistently high conductance without a clear annual signal, we used the average annual layer thickness of the 10 m above and below this section to determine the number of years within it. There are periods of detectable annual variations within this depth range, and they have approximately the same annual-layer thickness as the 10-m averages. A 10% uncertainty was assumed.

We assess the accuracy of WDC06A-7 by comparing it with two high-precision timescales: GICC05 and a new speleothem timescale from Hulu Cave. Because the age of the gas at a given depth is less than that of the ice surrounding it, we first need to calculate the age offset (Δage). We use the inferred accumulation rates and surface temperatures estimated from the $\delta^{18}O$ record constrained by the borehole temperature profile (Supplementary Information) in a steady-state firn-densification model[36]. The model is well-suited to WDC because it was developed using data from modern ice-core sites that span the full range of past WDC temperatures and accumulation rates. We calculate Δage using 200-yr smoothed histories of surface temperature and accumulation rate, a surface density of 370 kg m^{-3} and a close-off density of 810 kg m^{-3} (Supplementary Fig. 5a). The calculated present-day Δage is 210 yr, which is similar to the value, 205 yr, measured for WDC[37]. The steady-state model is acceptable for WDC because the surface temperature and accumulation rate vary more slowly than in Greenland. Because our primary purpose is to assess the accuracy of the WDC06A-7 timescale, calculation of Δage to better than a few decades is not necessary. The Δage uncertainty between 15 and 11 kyr ago is estimated to be 100 yr. The Δage uncertainty is estimated to be 150 yr for times before 20 kyr ago because of the colder temperatures and lower and less certain accumulation rates.

Because methane is well mixed in the atmosphere and should have identical features in both hemispheres, we use atmospheric methane measurements from WDC and the Greenland composite methane record[33] to compare WDC06A-7 and GICC05 at six times. The age differences are summarized in Supplementary Fig. 6 and the correlation and Δage uncertainties are shown in Supplementary Table 1. In Greenland, methane and $\delta^{18}O$ changes are nearly synchronous[38–40] and we therefore assume no Δage uncertainty in the Greenland gas timescale at times of abrupt change. An exception is at 24 kyr ago (Dansgaard–Oeschger event 2), when methane and $\delta^{18}O$ changes do not seem to be synchronous. We estimate the correlation uncertainty from the agreement of the methane records in Supplementary Fig. 5.

Speleothems can be radiometrically dated with U/Th and have smaller absolute age uncertainties than do annually resolved timescales in the glacial period[37]. Records of speleothem $\delta^{18}O$ show many abrupt changes that have been tied to the Greenland climate record[41,42]. However, the physical link between $\delta^{18}O$ variations in the caves and methane variations is not fully understood. Therefore, there is an additional and unknown correlation uncertainty in these comparisons. We compare WDC06A-7 with the new record from Hulu Cave, China, which is the best-dated speleothem record during this time interval. Comparisons can be made at only three times; our best estimate of the age differences is 100 yr or less.

The EDC timescale can be compared with the WDC06A-7 at a ~150-yr-long acid deposition event[15,16]. The two timescales agree within 100 yr, and we therefore do not adjust either timescale. The EDML timescale has been synchronized with the EDC timescale using sulphate matches[43]. The sulphate match that occurs during the 150-yr acid deposition event is marked in Fig. 2.

Sea-salt sodium measurements. Sea-salt sodium (ssNa) is the amount of Na that is of marine origin. The Na record was measured at the Trace Chemistry Laboratory at the Desert Research Institute. Na is one of many elements measured on the continuous-flow analysis system, which is coupled to two inductively coupled plasma mass spectrometers. The effective sampling resolution is ~1 cm. Details of the analytical set-up are described elsewhere[35,44–47]. Sea-salt Na is calculated assuming Na/Ca mass ratios of 26.3 for marine aerosols and 0.562 for average crust composition[48]. Sea-salt Na can be influenced by volcanic activity if the ratio of Na to Ca is different from the sea water and crustal ratios; the spike 20 kyr ago is part of an Na-rich but Ca-poor volcanic event. We present ssNa concentration in the main text instead of ssNa flux because wet deposition dominates at higher accumulation rates[49]. For comparison, the ssNa flux is shown in Supplementary Fig. 7.

Methane measurements. The methane concentration was measured in discrete samples at Oregon State University (OSU) and Pennsylvania State University (PSU) using automated melt–refreeze extraction and gas chromatography, with final concentration values reported on the NOAA04 concentration scale[50]. OSU data are corrected for gravitational fractionation, solubility and blanks as described in ref. 37. The gravitation fractionation correction assumes that $\delta^{15}N$ of N_2 is 0.3‰, a value based on late-Holocene measurements.

PSU methods were modelled on the basis of the OSU melt–refreeze system. The major difference between the OSU and PSU methods is the extraction cylinders; glass at OSU and stainless steel at PSU. Using stainless steel cylinders carries the

added problem of a blank associated with CH_4 outgassing, which we have estimated to be 19 ± 8 p.p.b. We have used a calculation similar to that derived in ref. 37, to estimate the amount of CH_4 left in the vessel after refreezing; we verified this using artificially degassed ice samples over which standard air was introduced and processed. These results indicate a 3.8% reduction in the measured headspace CH_4 value relative to the original trapped air, owing to solubility effects. The constant solubility and blank corrections were applied to all PSU data. In general, replicate samples from each depth were run on separate days to ensure that the final averaged data were not aliased by day-to-day instrument drifts. The average difference between replicate analyses of 1,316 individual depths run over 4 yr was 7 ± 8 p.p.b. (1σ). Finally, the PSU data were also corrected for gravitational fractionation by assuming that $\delta^{15}N$ of N_2 is 0.3‰ throughout.

To ensure that the PSU and OSU CH_4 data sets can be accurately merged into a single record, we performed an inter-calibration exercise involving a 100-m section of the WDC06A core (400–500 m) where both labs sampled for CH_4 every 2 m. By interpolating the OSU data to compare with the PSU data, we determined the average difference between the two labs over this 100 m interval to be 0.2 ± 9.9 p.p.b. (1σ). This result implies that we can merge CH_4 data from the two labs without correcting for inter-laboratory offsets.

Wilcoxon rank-sum test. Initial inspection of the WDC isotope record showed that warming was pulsed. We applied a sliding Wilcoxon rank-sum statistical test[51] to identify periods of significant change. A figure of the *P* values, for each individual Wilcoxon rank-sum test, is shown in Supplementary Fig. 8. A dashed line indicates the effective critical *P* value. Insignificant *P* values are plotted in grey, and significant *P* values are plotted in colours that correspond to timespan (block size) as in Fig. 2. The Wilcoxon rank-sum test makes no assumption of normality within the data and has been shown to be robust when used in windowing algorithms for the identification of periods of significant change in climate data[52]. Our windowing algorithm can also be applied using the more common Student's *t*-test. Though parametric, such an implementation has the benefit of a well-established method for correcting the degrees of freedom for autocorrelation within the data[53]. Applying either statistical test, we identify nearly identical periods of significant change in the data sets.

Climate modelling. To assess the effects of changing sea-ice conditions on precipitation-weighted $\delta^{18}O$ in Antarctica, we used the ECHAM4.6 climate model[24], implemented with the water isotope module[54]. Model simulations used a horizontal resolution of T42 (2.8° latitude by 2.8° longitude) with 19 vertical levels. The ECHAM4.6 model has been shown to reproduce Antarctic conditions realistically in the modern climate[13,55]. We used the sea surface temperatures from the PMIP2 fully coupled model experiments[25] for LGM conditions ~21 kyr ago. Those sea surface temperatures are prescribed as a model boundary condition for the atmospheric model runs with ECHAM4.6. We used a modern Antarctic ice-sheet configuration because the LGM configuration remains poorly known.

Model experiments were designed to test the sensitivity of $\delta^{18}O$ to changes in sea-ice extent. In the control experiment, sea ice forms at −1.7 °C and the model grid cell is set to 100% concentration below this threshold. The latitude of sea-ice coverage is decreased by lowering the ocean surface temperature threshold at which sea ice forms in the model. For the run with decreased sea ice, the freezing point was lowered from −1.7 to −3.7 °C. The amount of sea-ice reduction is not zonally uniform around Antarctica because of asymmetric gradients in the prescribed sea surface temperature. We note that model sea surface temperatures do not change whether model sea ice is present or not. Newly formed open water in the run with reduced sea ice is below the freezing point.

Integrated insolation. We calculate integrated annual insolation at latitude 65° S following the tables prepared in ref. 26. We also calculate integrated 'summer' and 'winter' insolation using a cut-off of 275 W m^{-2} (ref. 26; Supplementary Fig. 10).

32. Crosson, E. R. A cavity ring-down analyzer for measuring atmospheric levels of methane, carbon dioxide, and water vapor. *Appl. Phys. B* **92,** 403–408 (2008).
33. Dahl-Jensen, D., Gundestrup, N., Gogineni, S. P. & Miller, H. Basal melt at NorthGRIP modeled from borehole, ice-core and radio-echo sounder observations. *Ann. Glaciol.* **37,** 207–212 (2003).
34. Dansgaard, W. & Johnsen, S. J. A flow model and a time scale for the ice core from Camp Century, Greenland. *J. Glaciol.* **8,** 215–223 (1969).
35. Sigl, M. *et al.* A new bipolar ice core record of volcanism from WAIS Divide and NEEM and implications for climate forcing of the last 2000 years. *J. Geophys. Res.* **18,** 1151–1169 (2013).
36. Herron, M. M. & Langway, C. C. Firn densification: an empirical model. *J. Glaciol.* **25,** 373–385 (1980).
37. Mitchell, L. E., Brook, E. J., Sowers, T., McConnell, J. R. & Taylor, K. Multidecadal variability of atmospheric methane, 1000-1800 CE. *J. Geophys. Res.* **116,** G02007 (2011).
38. Huber, C. *et al.* Evidence for molecular size dependent gas fractionation in firn air derived from noble gases, oxygen, and nitrogen measurements. *Earth Planet. Sci. Lett.* **243,** 61–73 (2006).
39. Kobashi, T., Severinghaus, J. P., Brook, E. J., Barnola, J. M. & Grachev, A. M. Precise timing and characterization of abrupt climate change 8200 years ago from air trapped in polar ice. *Quat. Sci. Rev.* **26,** 1212–1222 (2007).
40. Severinghaus, J. P., Sowers, T., Brook, E. J., Alley, R. B. & Bender, M. L. Timing of abrupt climate change at the end of the Younger Dryas interval from thermally fractionated gases in polar ice. *Nature* **391,** 141–146 (1998).
41. Fleitmann, D. *et al.* Timing and climatic impact of Greenland interstadials recorded in stalagmites from northern Turkey. *Geophys. Res. Lett.* **36,** L19707 (2009).
42. Cheng, H. *et al.* Ice age terminations. *Science* **326,** 248–252 (2009).
43. Ruth, U. *et al.* "EDML1": a chronology for the EPICA deep ice core from Dronning Maud Land, Antarctica, over the last 150,000 years. *Clim. Past* **3,** 475–484 (2007).
44. Bisiaux, M. M. *et al.* Changes in black carbon deposition to Antarctica from two high-resolution ice core records, 1850-2000 AD. *Atmos. Chem. Phys.* **12,** 4107–4115 (2012).
45. McConnell, J. R. Continuous ice-core chemical analyses using inductively coupled plasma mass spectrometry. *Environ. Sci. Technol.* **36,** 7–11 (2002).
46. McConnell, J. R. *et al.* 20th-century industrial black carbon emissions altered arctic climate forcing. *Science* **317,** 1381–1384 (2007).
47. Pasteris, D. R., McConnell, J. R. & Edwards, R. High-resolution, continuous method for measurement of acidity in ice cores. *Environ. Sci. Technol.* **46,** 1659–1666 (2012).
48. Röthlisberger, R., Crosta, X., Abram, N. J., Armand, L. & Wolff, E. W. Potential and limitations of marine and ice core sea ice proxies: an example from the Indian Ocean sector. *Quat. Sci. Rev.* **29,** 296–302 (2010).
49. Alley, R. B. *et al.* Changes in continental and sea-salt atmospheric loadings in central Greenland during the most recent deglaciation: model-based estimates. *J. Glaciol.* **41,** 503–514 (1995).
50. Dlugokencky, E. J. *et al.* Conversion of NOAA atmospheric dry air CH_4 mole fractions to a gravimetrically prepared standard scale. *J. Geophys. Res.* **110,** D18306 (2005).
51. Wilcoxon, F. Individual comparisons by ranking methods. *Biom. Bull.* **1,** 80–83 (1945).
52. Mauget, S. A. Intra- to multidecadal climate variability over the continental United States: 1932-99. *J. Clim.* **16,** 2215–2231 (2003).
53. Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M. & Blade, I. The effective number of spatial degrees of freedom of a time-varying field. *J. Clim.* **12,** 1990–2009 (1999).
54. Hoffmann, G., Werner, M. & Heimann, M. Water isotope module of the ECHAM atmospheric general circulation model: a study on timescales from days to several years. *J. Geophys. Res.* **103,** 16871–16896 (1998).
55. Ding, Q. H., Steig, E. J., Battisti, D. S. & Kuttel, M. Winter warming in West Antarctica caused by central tropical Pacific warming. *Nature Geosci.* **4,** 398–403 (2011).

doi:10.1038/nature12336

Digit loss in archosaur evolution and the interplay between selection and constraints

Merijn A. G. de Bakker[1], Donald A. Fowler[1]†, Kelly den Oude[1], Esther M. Dondorp[1]†, M. Carmen Garrido Navas[1]†, Jaroslaw O. Horbanczuk[2], Jean-Yves Sire[3], Danuta Szczerbińska[4] & Michael K. Richardson[1]

Evolution involves interplay between natural selection and developmental constraints[1–3]. This is seen, for example, when digits are lost from the limbs during evolution[1,3,4]. Extant archosaurs (crocodiles and birds) show several instances of digit loss[3,5,6] under different selective regimes, and show limbs with one, two, three, four or the ancestral number of five digits. The 'lost' digits sometimes persist for millions of years as developmental vestiges[7–10]. Here we examine digit loss in the Nile crocodile and five birds, using markers of three successive stages of digit development. In two independent lineages under different selection, wing digit I and all its markers disappear. In contrast, hindlimb digit V persists in all species sampled, both as cartilage, and as *Sox9*- expressing precartilage domains, 250 million years after the adult digit disappeared. There is therefore a mismatch between evolution of the embryonic and adult phenotypes. All limbs, regardless of digit number, showed similar expression of sonic hedgehog (*Shh*). Even in the one-fingered emu wing, expression of posterior genes *Hoxd11* and *Hoxd12* was conserved, whereas expression of anterior genes *Gli3* and *Alx4* was not. We suggest that the persistence of digit V in the embryo may reflect constraints, particularly the conserved posterior gene networks associated with the zone of polarizing activity (ZPA[11]). The more rapid and complete disappearance of digit I may reflect its ZPA-independent specification, and hence, weaker developmental constraints. Interacting with these constraints are selection pressures for limb functions such as flying and perching. This model may help to explain the diverse patterns of digit loss in tetrapods. Our study may also help to understand how selection on adults leads to changes in development.

Digit loss is defined as the complete loss of all phalanges and the metapodial bone; it should be distinguished from digit reduction, in which only phalanges are lost[3]. Digit loss can be adaptive. It reduces the mass of the distal limb, and therefore its moment of inertia; this conserves energy during running and flying[12–14]. As a result, there is directional selection for digit loss in the wing of all birds studied here, and in the flying ancestors of the ratites[15]. Hindlimb digit loss is similarly an adaptation for running, and is seen in ratites, ungulates and some dinosaurs.

In the flightless ostrich, the wings are miniaturized, but still have three digits, presumably because they are used for sexual display, to help in turning (banking) while running, and to shade the eggs. The emu wing has none of these functions[16,17] and has lost digits because of relaxation of stabilizing selection through disuse, so that changes in the wing are selectively neutral[16,17]. Selection pressures may favour the retention of digits. This may be the case with hindlimb digit I in many Neoaves. This digit is partly reduced, having only a vestigial metatarsal I (Figs 1 and 2)[3,6], nevertheless, its distal part is developed fully, presumably because it is important for perching[18].

Given these varying selection pressures, the patterns of digit reduction in tetrapod evolution are complex, making it difficult to identify consistent trends or rules (Supplementary Fig. 1). For example, in placental mammals with digit loss, digit I is usually (but not always) the first to disappear (Supplementary Fig. 1). In pigs, digits III and IV are the main functional elements; digits II and V are smaller and digit I is vestigial[19]. Lizards frequently lose digit I in evolution, but not digit V (although the latter frequently loses phalanges[3]). Turtles and tortoises may lose digit I from the hand, but not from the foot[3]. In the hindlimb of archosaurs, digit V is typically reduced or lost.

Species differences in digit loss are also seen after chemical treatment of embryos. When such treatment entirely eliminates a digit from the foot, it is commonly digit I in anurans but digit V in urodeles[1,4]. In

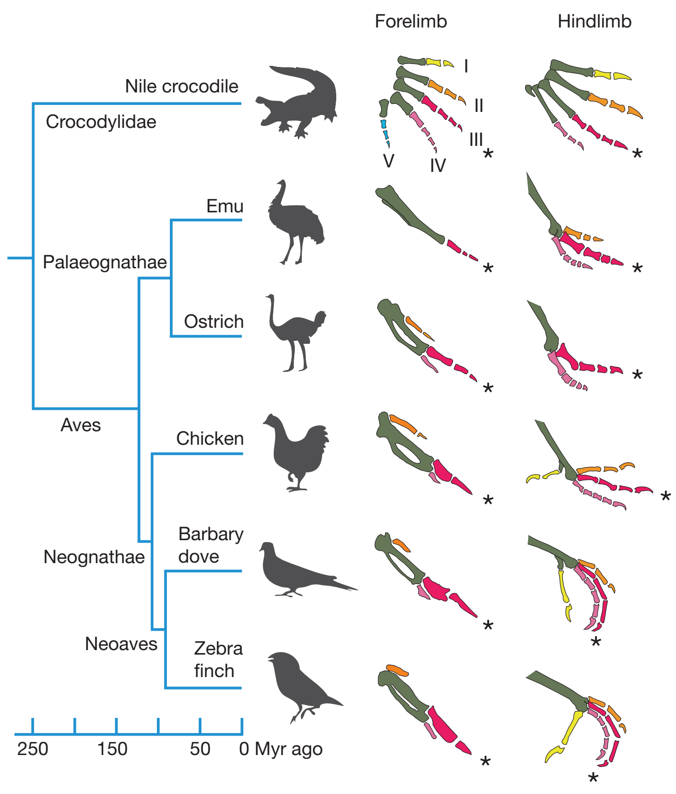


Figure 1 | Changes in adult digit number across archosaur phylogeny. Phylogeny and divergence estimates based on refs 15, 21. Schematic drawings of the fore- and hindlimb of the studied species, made from original museum specimens (Supplementary Table 3). The number of digits in the adult skeleton varies from five (in the hindlimb of the crocodile) to one (in the wing of the emu). Yellow, digit I; orange, digit II; red, digit III; purple, digit IV; blue, digit V. Myr ago, million years ago. Asterisks indicate digit III.

[1]Department of Integrative Zoology, Institute of Biology, Leiden University, Sylvius Laboratory, Sylviusweg 72, 2333BE Leiden, the Netherlands. [2]Institute of Genetics and Animal Breeding, Polish Academy of Sciences, 05-552 Jastrzębiec, Poland. [3]Evolution & Développement du squelette, UMR 7138, Université Pierre et Marie Curie, 7 Quai Saint-Bernard, Bat A2, Case 5, 75005 Paris, France. [4]Department of Poultry and Ornamental Bird Breeding, Western Pomeranian University of Technology, 20 Judyma Street, 71-466 Szczecin, Poland. †Present addresses: Redpath Museum, McGill University, 859 Rue Sherbrooke Ouest, Montreal, Quebec H3A 2K6, Canada and Department of Biology, McGill University, Stewart Biology Building, 1205 Docteur Penfield Montreal, Quebec H3A 1B1, Canada (D.A.F.); Naturalis Biodiversity Center, PO Box 9517, 2300 RA Leiden, The Netherlands (E.M.D.); Department of Genetics, University of Leicester, Adrian Building, University Road, Leicester LE1 7RH, UK (M.C.G.-N.).

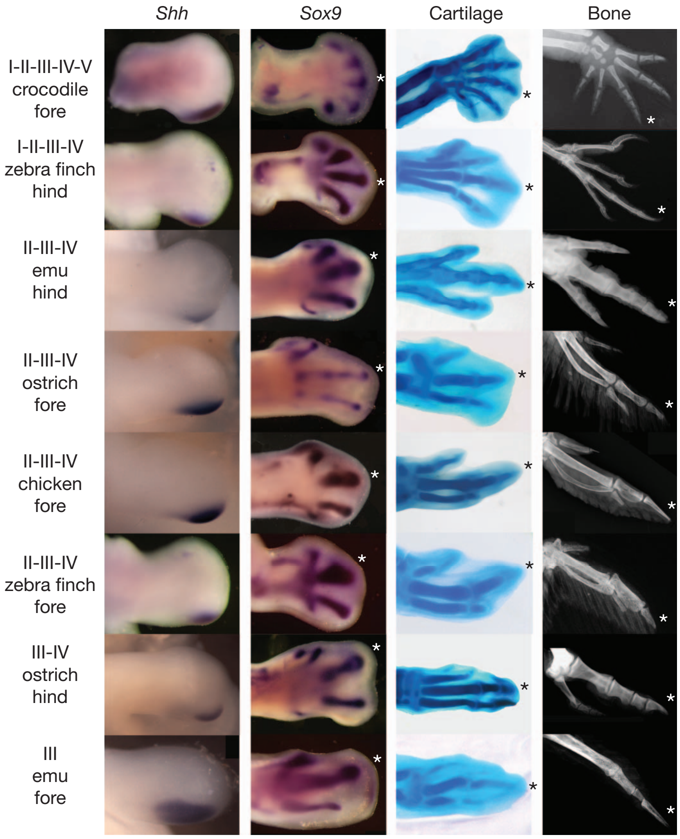


Figure 2 | Comparison of developmental and adult limb phenotypes in archosaurs. *Shh* expression at early digital plate stage, *Sox9* expression of the early prechondrocyte marker *Sox9* (refs 9, 20) at late digital plate (paddle) stage; cartilage, Alcian blue staining for cartilage at the digital ray stage; and bone, digital radiographs of adult skeletons from recently prepared specimens in which the soft tissue has not been removed. In all pictures, anterior is to the top and distal to the right. Asterisks indicate digit III.

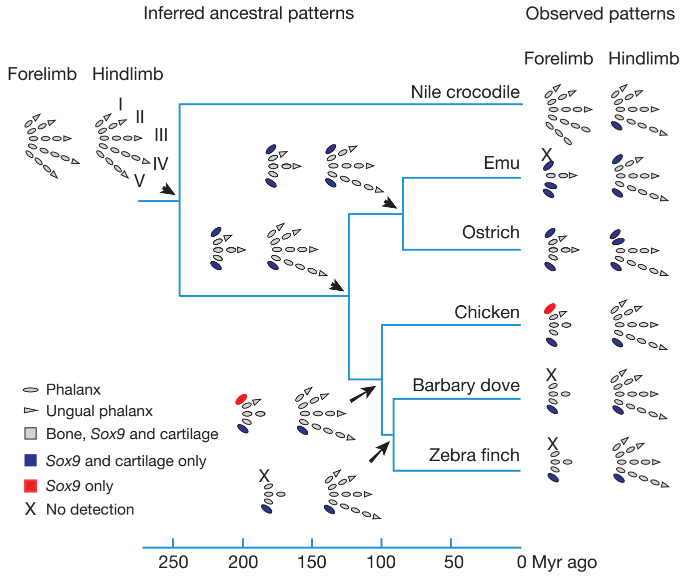


Figure 3 | Developmental sequences of digit loss mapped onto phylogeny, with inferred ancestral conditions. Right, schematic diagrams of forelimb and hindlimb pairs, showing the stage to which each digit was found to develop in this study. The colours indicate the final stage (*Sox9* domain, cartilage or bone) to which that digit developed. When none of these phenotypes was detected, the digit is indicated by X. Left, the inferred ancestral states at all nodes in the cladogram (based on refs 15, 21).

both cases, the digit most sensitive to chemical treatment tends to be the digit developing last in the embryo, and the digit that disappears first in evolution[4]. These findings led to the idea that digit loss, as with many other morphological patterns, may be determined by a combination of developmental constraints and natural selection[3,4].

The archosaur species studied here provide examples of limbs possessing one, two, three, four and five digits (Figs 1 and 2). The study species are the Nile crocodile (*Crocodylus niloticus*), emu (*Dromaius novaehollandiae*), ostrich (*Struthio camelus*), chicken (*Gallus gallus*), Barbary dove (*Streptopelia risoria*) and zebra finch (*Taeniopygia guttata*). The crocodilian forelimb is the only limb in our sample that has all five adult digits (Figs 1 and 2)[5]. It also has five cartilaginous digits in the embryo, each preceded by an expression domain of *Sox9* (Fig. 2), an early marker of prechondrocyte condensations[20].

Digit V has been lost from the adult skeleton in all other limbs sampled. Despite this, all limbs develop a transient *Sox9* primordium in the digit V position (Fig. 2 and Supplementary Fig. 2) and all limb digit V domains progress to the cartilage stage (Fig. 2 and Supplementary Fig. 3). Therefore, considering the evolutionary timescale (Fig. 1)[15,21], the cartilage-forming pathways for digit V have persisted at least 250 million years after digit V disappeared from the archosaurs' adult hindlimb (Fig. 3).

Digit I is lost in the forelimbs of all adult birds studied, and in the hindlimbs of the emu and ostrich. It is represented only by weak cartilage staining in the ostrich wing[8,10], and only by *Sox9* and peanut agglutinin (PNA) expression in the chicken wing[7,9]; no *Sox9* or cartilage develops in the digit I position in the wings of the emu or Neoaves (Barbary dove and zebra finch; Fig. 2 and Supplementary Figs 2 and 3).

Therefore, on the basis of the markers used here, and considering the estimated age of the clade Aves[15,21], digit I and its underling pathways are no longer detectable after 120 million years (Fig. 3).

Digit loss is carried furthest in the ostrich and emu. The ostrich hindlimb has only two digits in the adult skeleton, but retains a full five *Sox9* expression domains in the embryo, followed by five chondrified condensations (Fig. 2)[8,10]. The wing in the adult emu has only a single digit, but in development shows distinct *Sox9* domains for digits II, III and IV and a weak digit V domain; this four-digit pattern is seen also in the cartilaginous skeleton (Fig. 2)[17].

These data show three things: first, expression of the early cartilage marker *Sox9* can persist even when multiple digits have been lost; second, there are two independent lineages in which the *Sox9* domain for digit I has disappeared, whereas that for digit V has persisted; and third, there is a mismatch between the adult and embryonic phenotypes such that the pattern of digits in the adult skeleton is not congruent with the pattern of digit primordia in the embryo.

One mechanism proposed[22–24] to control digit number is based on changes in the timing of SHH expression at the posterior margin of the limb bud, in the zone of polarizing activity (ZPA)[11]. The ZPA–SHH pathway is deeply conserved in evolution[25] and modulates the expression of posterior *Hox* genes, as well as anterior patterning genes such as *Gli3* and *Alx4*. These in turn help to establish the digit pattern (possibly through a Turing-type pre-pattern mechanism[26]). Loss of SHH or ZPA function in the mouse or chicken[27] leads to a malformed limb with a single, reduced digit. Furthermore, experimental inhibition of progressively earlier phases of SHH signalling causes correspondingly fewer digits to develop[22,23]. Finally, in the lizard genus *Hemiergis*, species with fewer digits seem to show earlier termination of *Shh* expression during limb development[24].

Taken together, these recent studies may lead us to expect significant differences in the timing of *Shh* expression among our sample of archosaurs. However, we find that *Shh* remains expressed at least until the digits are already established as *Sox9*-expressing primordia, regardless of the number of adult digits in the limb (Fig. 2 and Supplementary Figs 2 and 4).

In addition to conserved patterns of posterior *Shh* expression, we also find conservation in the early posterior expression domains of

Hoxd11 and *Hoxd12* (Fig. 4a). Even the emu wing, which is vestigial, miniaturized and has only one digit, shows normal early expression domains of the posterior genes (Fig. 4a). By contrast, the early expression patterns of the anterior genes *Gli3* and *Alx4* show more evolutionary variation (Fig. 4a). This suggests that the posterior pathways are more constrained than anterior ones in our sample.

We suggest that the conserved patterns of gene expression at the posterior margin of the limb (Fig. 2 and Supplementary Figs 2 and 4) could explain the persistence of digit V primordia in embryos. The ZPA is pivotal for the development of the limb and influences the phenotype of all digits except digit I[22,23]. We therefore suggest that digit I is under weaker developmental constraints than digit V. This interpretation is strengthened by our observation (Fig. 3) of the parallel disappearance of digit I pathways in two species of Neoaves, and in the emu. When the conserved SHH signalling comes to an end (stages 26 and 27 in the hindlimb and forelimb, respectively; Supplementary Fig. 4b), we suggest that it becomes easier to supress the growth and differentiation of digit V. Interestingly, the general evolutionary trend we find towards reduction in digit I markers (Fig. 3) is consistent with

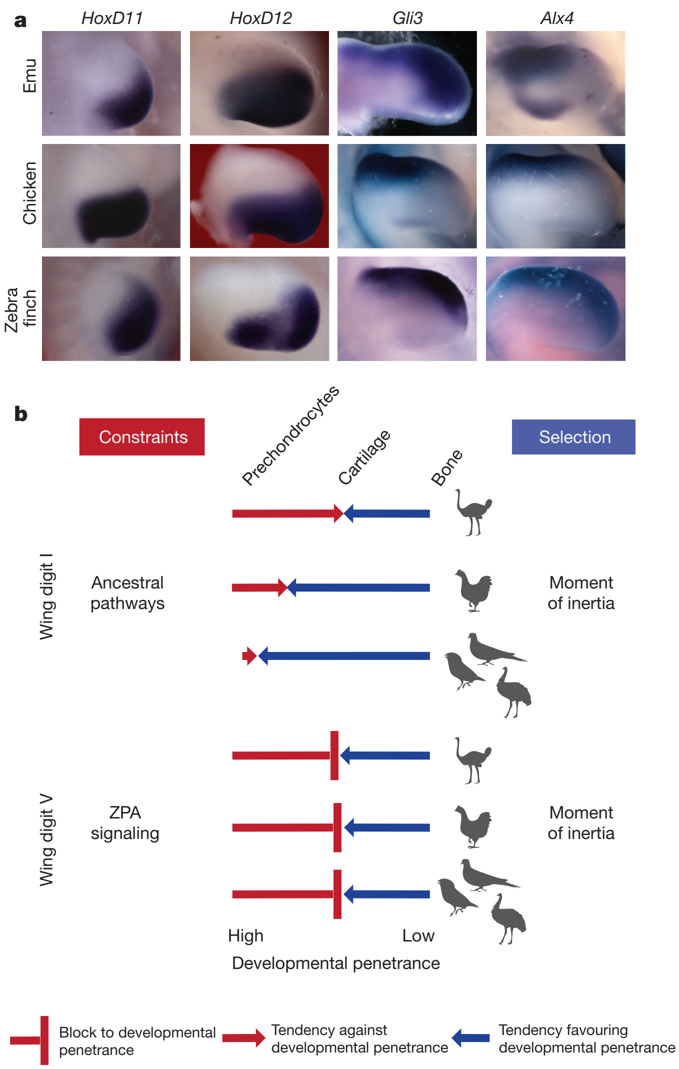


Figure 4 | Constraints versus selection pressures in the forelimb. a, The two *Hoxd* genes show a similar posterior expression in all species. The anterior expressed *Gli3* and *Alx4* show modified expression in the emu and zebra finch. Thus, *Gli3* is much more extensively expressed in the autopod, as is *Alx4*, and it shows an extra posterior domain in the emu wing. **b**, The balancing act between selection and developmental constraints; a model of our hypothesis to explain the different rates at which digit I and V pathways disappear in evolution. See Supplementary Fig. 5 for hindlimbs.

previous studies showing a gradual reduction of wing digits in avian evolution, as indicated by gradual loss of claws (reviewed in ref. 6).

Our results are consistent with a model (Fig. 4b) in which an interplay between natural selection and developmental constraints gives very different outcomes for the development of digit I and digit V in different lineages. A balancing act between selection and constraints may help to account for digit loss in other taxa. In land tortoises, for example, if a digit is lost entirely in one step, it is always digit I[28].

More generally, our model may help in understanding why the embryo digit phenotype is not always congruent with the adult digit pattern. For example, we show that the same selection pressure for digit loss has different degrees of developmental penetrance in wing digit I and V (Fig. 4b). The 'developmental penetrance' of a change in adult phenotype is the degree to which early developmental pathways are modified in order to produce that change[29]. Loss of adult digit I in the wing has a high developmental penetrance because constraints are weaker; it is therefore easier to modify developmental pathways (Fig. 4b) and lose digit I markers completely. The converse is true of digit V. In summary, the interplay between selection and constraints may be important for shaping not only adult phenotypes, but embryo phenotypes as well.

METHODS SUMMARY

The eggs of the Barbary dove (*Streptopelia risoria*) and the zebra finch (*Taeniopygia guttata*) were a gift of C. J. ten Cate and K. Riebel. The chicken (*Gallus gallus*), ostrich (*Struthio camelus*) and emu (*Dromaius novaehollandiae*) eggs were obtained from commercial breeders or from our own colony. The obtained eggs, embryos and radiographs of the Nile crocodile (*Crocodylus niloticus*) were a gift of S. Martin from 'La Ferme aux Crocodiles' in Pierrelatte (France). In total, 260 embryos were photographed and analysed (see Supplementary Table 1). The hybridization protocols were as described previously[9].

The chicken *Shh* plasmid was a gift from C. J. Tabin, and the *Alx4* from M. Schwarz, Salk Institute, San Diego. The probes for *Sox9, Gli3, HoxD11 and Hoxd12* were made in-house and their sequences deposited at the National Center for Biotechnology Information (NCBI) (Supplementary Table 2 and Supplementary Fig. 6). Embryos were grouped into stages according to the staging tables of Hamburger and Hamilton[30]. When in doubt, the decision was made based on the hindlimb characters as given in ref. 30. The radiographs were made of fresh specimens. The skeleton preparations used for the schematic drawings in Fig. 1 are of museum specimens or of our own collection (Supplementary Table 3).

Full Methods and any associated references are available in the online version of the paper.

Received 9 January; accepted 24 May 2013.
Published online 7 July 2013.

1. Alberch, P. & Gale, E. A. Size dependence during the development of the amphibian foot: colchicine-induced digital loss and reduction. *J. Embryol. Exp. Morphol.* **76,** 177–197 (1983).
2. Richardson, M. K. & Chipman, A. D. Developmental constraints in a comparative framework: a test case using variations in phalanx number during amniote evolution. *J. Exp. Zool. B Mol. Dev. Evol.* **296,** 8–22 (2003).
3. Shapiro, M. D., Shubin, N. H. & Downs, J. P. In *Fins Into Limbs: Evolution, Development and Transformation* (ed. Hall, B. K.) 225–244 (Univ. of Chicago Press, 2007).
4. Alberch, P. & Gale, E. A. A developmental analysis of an evolutionary trend: digital reduction in amphibians. *Evolution* **39,** 8–23 (1985).
5. Müller, G. B. & Alberch, P. Ontogeny of the limb skeleton in *Alligator mississippiensis*: developmental invariance and change in the evolution of archosaur limbs. *J. Morphol.* **203,** 151–164 (1990).
6. Richardson, M. K. in *From Clone to Bone: The Synergy of Morphological and Molecular Tools in Palaeobiology* 328–362 (eds Asher, R. J. & Müller, J.) (Cambridge Univ. Press, 2012).
7. Larsson, H. C. & Wagner, G. P. Pentadactyl ground state of the avian wing. *J. Exp. Zool. B Mol. Dev. Evol.* **294,** 146–151 (2002).
8. Feduccia, A. & Nowicki, J. The hand of birds revealed by early ostrich embryos. *Naturwissenschaften* **89,** 391–393 (2002).
9. Welten, M. C., Verbeek, F. J., Meijer, A. H. & Richardson, M. K. Gene expression and digit homology in the chicken embryo wing. *Evol. Dev.* **7,** 18–28 (2005).
10. Kundrát, M. Primary chondrification foci in the wing basipodium of *Struthio camelus* with comments on interpretation of autopodial elements in Crocodilia and Aves. *J. Exp. Zoolog. B Mol. Dev. Evol.* **312,** 30–41 (2009).
11. Harfe, B. D. Keeping up with the zone of polarizing activity: new roles for an old signaling center. *Dev. Dyn.* **240,** 915–919 (2011).

12. Abourachid, A. & Renous, S. Bipedal locomotion in ratites (Paleognatiform): examples of cursorial birds. *Ibis* **142,** 538–549 (2000).
13. Schaller, N. U., D'Aout, K., Villa, R., Herkner, B. & Aerts, P. Toe function and dynamic pressure distribution in ostrich locomotion. *J. Exp. Biol.* **214,** 1123–1130 (2011).
14. Van den berg, C. & Rayner, J. M. V. The moment of inertia of bird wings and the inertial power requirement for flapping flight. *J. Exp. Biol.* **198,** 1655–1664 (1995).
15. Haddrath, O. & Baker, A. J. Multiple nuclear genes and retroposons support vicariance and dispersal of the palaeognaths, and an Early Cretaceous origin of modern birds. *Proc. R. Soc. B* **279,** 4617–4625 (2012).
16. Maxwell, E. E. & Larsson, H. C. Osteology and myology of the wing of the Emu (*Dromaius novaehollandiae*), and its bearing on the evolution of vestigial structures. *J. Morphol.* **268,** 423–441 (2007).
17. Nagai, H. *et al.* Embryonic development of the Emu, *Dromaius novaehollandiae. Dev. Dyn.* **240,** 162–175 (2011).
18. Middleton, K. M. The morphological basis of hallucal orientation in extant birds. *J. Morphol.* **250,** 51–60 (2001).
19. Sears, K. E. *et al.* Developmental basis of mammalian digit reduction: a case study in pigs. *Evol. Dev.* **13,** 533–541 (2011).
20. Lorda-Diez, C. I., Montero, J. A., Diaz-Mendoza, M. J., Garcia-Porrero, J. A. & Hurle, J. M. Defining the earliest transcriptional steps of chondrogenic progenitor specification during the formation of the digits in the embryonic limb. *PLoS ONE* **6,** e24546 (2011).
21. Hugall, A. F., Foster, R. & Lee, M. S. Y. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Syst. Biol.* **56,** 543–563 (2007).
22. Scherz, P. J., McGlinn, E., Nissim, S. & Tabin, C. J. Extended exposure to Sonic hedgehog is required for patterning the posterior digits of the vertebrate limb. *Dev. Biol.* **308,** 343–354 (2007).
23. Zhu, J. *et al.* Uncoupling Sonic hedgehog control of pattern and expansion of the developing limb bud. *Dev. Cell* **14,** 624–632 (2008).
24. Shapiro, M. D., Hanken, J. & Rosenthal, N. Developmental basis of evolutionary digit loss in the Australian lizard Hemiergis. *J. Exp. Zool. B Mol. Dev. Evol.* **297,** 48–56 (2003).
25. Gillis, J. A., Dahn, R. D. & Shubin, N. H. Shared developmental mechanisms pattern the vertebrate gill arch and paired fin skeletons. *Proc. Natl Acad. Sci. USA* **106,** 5720–5724 (2009).
26. Sheth, R. *et al.* Hox genes regulate digit patterning by controlling the wavelength of a Turing-type mechanism. *Science* **338,** 1476–1480 (2012).
27. Pagan, S. M., Ros, M. A., Tabin, C. & Fallon, J. F. Surgical removal of limb bud Sonic hedgehog results in posterior skeletal defects. *Dev. Biol.* **180,** 35–40 (1996).
28. Crumly, C. R. & Sánchez-Villagra, M. R. Patterns of variation in the phalangeal formulae of land tortoises (Testudinidae): developmental constraint, size, and phylogenetic history. *J. Exp. Zool. B Mol. Dev. Evol.* **302,** 134–146 (2004).
29. Richardson, M. K. Vertebrate evolution: the developmental origins of adult variation. *Bioessays* **21,** 604–613 (1999).
30. Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. *J. Morphol.* **88,** 49–92 (1951).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. C. M. Welten for technical advice; W. Bruins for help with collecting the emu embryos; D. van der Marel for taking the radiographs; H. J. Meijer for advice on the phylogeny and for guiding D.A.F. through the the collections at the Smithsonian Institution; the Delaware Museum of Natural History for allowing D.A.F. to measure their specimens; and P. den Hartog, H. Koolmoes and S. de Schaaf-Timmerman for collecting the zebra finch and Barbary dove eggs.

Author Contributions M.A.G.d.B. conceived the research, wrote the paper, carried out probe design and synthesis, embryo harvesting, gene-expression studies and analysis, and created the figures; D.A.F. carried out gene expression studies, embryo harvesting, and created the adult skeleton figures; K.d.O. carried out gene expression studies and embryo collection; E.M.D. carried out gene-expression studies and collected embryos; M.C.G.-N. carried out alcian blue staining and collected embryos; J.O.H. carried out emu and ostrich embryo collection; D.S. carried out emu embryo collection and incubation; J.-Y.S. carried out crocodile embryo collection and incubation; and M.K.R. conceived the research, wrote the paper, carried out embryo collection and analysis, created figures and provided funding and facilities.

Author Information New sequences of probes used here are deposited in GenBank under accession numbers KC667071 (chicken *Gli3*), JF317542 (chicken *HoxD11*), JF317555 (chicken *HoxD12*), JQ717196 (Nile crocodile *Sox9*) and JQ717195 (emu *Sox9*). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.K.R. (m.k.richardson@biology.leidenuniv.nl).

METHODS

Incubation of eggs and collection of embryos. The chicken and ostrich eggs were from commercial breeders, as were some of the emu eggs. The Barbary dove and zebra finch eggs were a kind gift of C. J. ten Cate and K. Riebel. All chicken, dove and zebra finch eggs were incubated and collected in Leiden as were some of the ostrich, emu and crocodile eggs. The remaining emu and ostrich eggs were incubated and collected in Poland. Some of the crocodile eggs donated by 'La Ferme aux Crocodiles' were collected on site in Pierrelatte, France.

For incubation temperatures and range of harvested stages see Supplementary Table 1. After candling the eggs, embryos were removed and transferred to ice-cold PBS in a Petri dish. The amnion was removed, and the embryo staged and fixed in ice-cold 4% paraformaldehyde in PBS at 4 °C overnight. Embryos were grouped into stages according to the staging tables described previously[30]. When in doubt, the decision was made based on the hindlimb characters as given in ref. 30. The next day they were dehydrated in a graded methanol series and stored in 100% methanol at −20 °C.

Probe synthesis. For the probes synthesized in-house we isolated total RNA from one embryo with Trizol (Invitrogen) and reverse transcribed it with SuperScript III (Invitrogen) or RevertAid (Fermantas). On these templates we performed polymerase chain reactions (PCRs) with specific primers and cloned the PCR products in the TOPOTA-PCRII vector (Invitrogen). The inserted PCR products were checked by Sanger sequencing. These results were deposited at NCBI (Supplementary Table 2) and compared with known sequences (Supplementary Fig. 6). From the sequence data we could also determine the restriction enzyme needed to linearize the isolated plasmid (Qiagen miniprep column) as the RNA polymerase for making the anti-sense RNA probe labelled with digoxigenin. For this we used Sp6 or T7 RNA-polymerase (Roche or Fermentas) and the digoxigenin RNA labelling mix (Roche). The probe was cleaned with RNeasy purification columns of Qiagen.

Alcian blue staining. Alcian blue stains acidic polysaccharides such as glycosaminoglycans in cartilages. The fixed embryos were washed in 70% ethanol followed by acid alcohol (70% ethanol + 1% concentrated hydrochloric acid), stained in 0.03% (w/v) alcian blue in acid alcohol and washed in acid alcohol.

Whole-mount *in situ* hybridization. The embryos were rehydrated through graded methanols, digested lightly with proteinase K (20 to 40 μg ml^{-1} PBS) for 20 min and re-fixed in paraformaldehyde in PBS after several washes in PBST (PBS pH 7.2 with 0.1% Tween-20). This was followed by a pre-hybridization step at 60 °C for at least 3 h or until the embryo had sunk. The hybridization mixture consisted of: 50% Formamide, 2% Boehringer blocking powder, 5× SSC (standard sodium citrate buffer, pH 7), 1 mg ml^{-1} total RNA, 50 μg ml^{-1} heparin, 0.1% Triton X-100, 0.1% CHAPS (3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate) and 5 mM EDTA. After the pre-hybridization mix was removed we added 400 ng ml^{-1} specific probe to fresh hybridization mixture preheated to 60 °C before adding it to the embryo; it was incubated in this mix at 60 °C overnight with shaking. The next day the specific probe mixture was removed, collected and stored at −20°C for reuse.

Several stringent washes were done at 60°C to remove non-specifically bound probe (2× SSC, 0.1% CHAPS, 50% formamide); (2× SSC, 0.1% CHAPS); (0.2× SSC, 0.1% CHAPS). After washing several times at room temperature (20 °C) with TBST (0.1 M tris buffered saline, pH 7.5, 0.1% Tween-20) the embryos were pre-incubated with 10% sheep serum in TBST for 90 min at room temperature followed by overnight incubation with sheep anti-digoxigenin conjugated to alkaline phosphatase (Roche; 1:5,000 dilution in 10% sheep serum in TBST at 4 °C overnight). The next day the non-specifically bound antibodies were washed away by several washes with TBST, including one overnight. The embryos were brought to a higher pH by washing them in NTT buffer (0.1 M sodium chloride, 0.1 M Tris-HCl, 0.1% Tween-20, pH 9.5). The enzyme reaction of alkaline phosphate with NBT/BCIP (nitro blue tetrazolium chloride/5-Bromo-4-chloro-3-indolyl phosphate) or BM purple (both Roche) as substrate results in a blue precipitate. The development of the stain was checked regularly and stopped by washing several times in TBST, removing the substrate and chromogens, and lowering the pH.

Radiography. All radiographs were made using fresh specimens. The radiograph of the zebra finch was made on film and later digitized at the Institute of Biology Leiden (IBL), the Netherlands. The chicken, emu and ostrich digital radiographs were made by D. van Marel of the Naturalis Biodiversity Center, the Netherlands. The crocodile was radiographed at 'La Ferme aux Crocodiles', France.

doi:10.1038/nature12438

Emergence of structural and dynamical properties of ecological mutualistic networks

Samir Suweis[1], Filippo Simini[2,3], Jayanth R. Banavar[4] & Amos Maritan[1]

Mutualistic networks are formed when the interactions between two classes of species are mutually beneficial. They are important examples of cooperation shaped by evolution. Mutualism between animals and plants has a key role in the organization of ecological communities[1–3]. Such networks in ecology have generally evolved a nested architecture[4,5] independent of species composition and latitude[6,7]; specialist species, with only few mutualistic links, tend to interact with a proper subset of the many mutualistic partners of any of the generalist species[1]. Despite sustained efforts[5,8–10] to explain observed network structure on the basis of community-level stability or persistence, such correlative studies have reached minimal consensus[11–13]. Here we show that nested interaction networks could emerge as a consequence of an optimization principle aimed at maximizing the species abundance in mutualistic communities. Using analytical and numerical approaches, we show that because of the mutualistic interactions, an increase in abundance of a given species results in a corresponding increase in the total number of individuals in the community, and also an increase in the nestedness of the interaction matrix. Indeed, the species abundances and the nestedness of the interaction matrix are correlated by a factor that depends on the strength of the mutualistic interactions. Nestedness and the observed spontaneous emergence of generalist and specialist species occur for several dynamical implementations of the variational principle under stationary conditions. Optimized networks, although remaining stable, tend to be less resilient than their counterparts with randomly assigned interactions. In particular, we show analytically that the abundance of the rarest species is linked directly to the resilience of the community. Our work provides a unifying framework for studying the emergent structural and dynamical properties of ecological mutualistic networks[2,5,10,14].

Statistical analyses of empirical mutualistic networks indicate that a hierarchical nested structure is prevalent and is characterized by nestedness values that are consistently higher than those found in randomly assembled networks with the same number of species and interactions[1,6]. Nevertheless, the degree of nestedness varies among networks. Recently[5,10], it has been argued that nestedness increases biodiversity and begets stability, but these results are in conflict with robust theoretical evidences showing that ecological communities with nested interactions are inherently less stable than unstructured ones[12,14,15] and that mutualism could be detrimental to persistence[11,15]. We aim to elucidate general optimization mechanisms underlying network structure and its influence on community dynamics and stability.

There is a venerable history of the use of variational principles for understanding nature, which has led to major advances in many subfields of physics, including classical mechanics, electromagnetism, relativity, and quantum mechanics. Our goal is to determine the appropriate variational principle that characterizes a mutualistic community in the absence of detailed knowledge of the nature and strengths of the interactions between species and their environment. We begin by showing that increases in the abundances of the species lead to an increase in the total number of individuals (henceforth referred to as the total population) within the mutualistic community. We then show that, under stationary conditions, the total population is directly correlated with nestedness and vice versa. Finally, we demonstrate that nested mutualistic communities are less resilient than communities in which species interact randomly. These results suggest a simple and general optimization principle: key aspects of mutualistic network structure and its dynamical properties could emerge as a consequence of the maximization of the species abundance in the mutualistic community (see Fig. 1).

We consider a community comprising a total of S interacting species (see Methods), in which population dynamics is driven by interspecific interactions. We model mutualistic and competitive species interactions using both the classical Holling type I and II functional responses[16–18] (Supplementary Information). We perform a controlled numerical experiment at the stable stationary state by holding fixed the number of species, the strengths of the interactions, and the connectance (the fraction of non-zero interactions), and seek to maximize individual species population abundances by varying the network architecture. The simplest approach consists of repeatedly rewiring the interactions of a randomly drawn species so as to increase its abundance, that is, each selected species attempts to change its mutualistic partners in order to enhance the benefit obtained from its interactions (see Methods and Supplementary Information). The optimization principle may then be interpreted within an adaptive evolutionary framework within which species maximize the efficiency of resource usage[19,20] and minimize their chances of becoming extinct owing to stochastic perturbations[21,22]. Interestingly, we find that enhancements in the abundance of any given species most often results in growth of the total population along with a concomitant increase of the nestedness (see Fig. 1). These results demonstrate the existence of a correlation between nestedness and species abundance and highlight a non-trivial collective effect through which each successful switch affects the abundances of all species, leading to an inexorable increase, on average, of the total number of individuals in the community.

In order to make analytical progress and to better understand the correlation found between the optimization of individual species abundances, the total number of individuals in the community and nestedness, we turn to a mean field approximation[5] in which the mutualistic (and competitive) interactions are assumed to have the same magnitude. Within this approximation, we are able to prove that (see Supplementary Information for mathematical details): (1) an increase in the abundance of any species more often than not leads to a net increase in the total population of the community; and (2) communities with larger total population have interaction matrices with higher nestedness and vice versa. The intraspecific (plant–plant and animal–animal) interactions have a key but secondary role compared to the mutualistic (plant–animal) interactions. The main effect of the intraspecific interactions is to break the degeneracy in the network overlap (Supplementary Information). Extensive numerical simulations in the more general, non-mean field case of heterogeneous interactions also confirm these findings. The

[1]Dipartimento di Fisica e Astronomia 'G. Galilei' & CNISM, INFN, Università di Padova, Via Marzolo 8, 35131 Padova, Italy. [2]Center for Complex Network Research and Department of Physics, Biology and Computer Science, Northeastern University, Boston, Massachusetts 02115, USA. [3]Institute of Physics, Budapest University of Technology and Economics Budafoki ut 8, Budapest H-1111, Hungary. [4]Department of Physics, University of Maryland, College Park, Maryland 20742, USA.

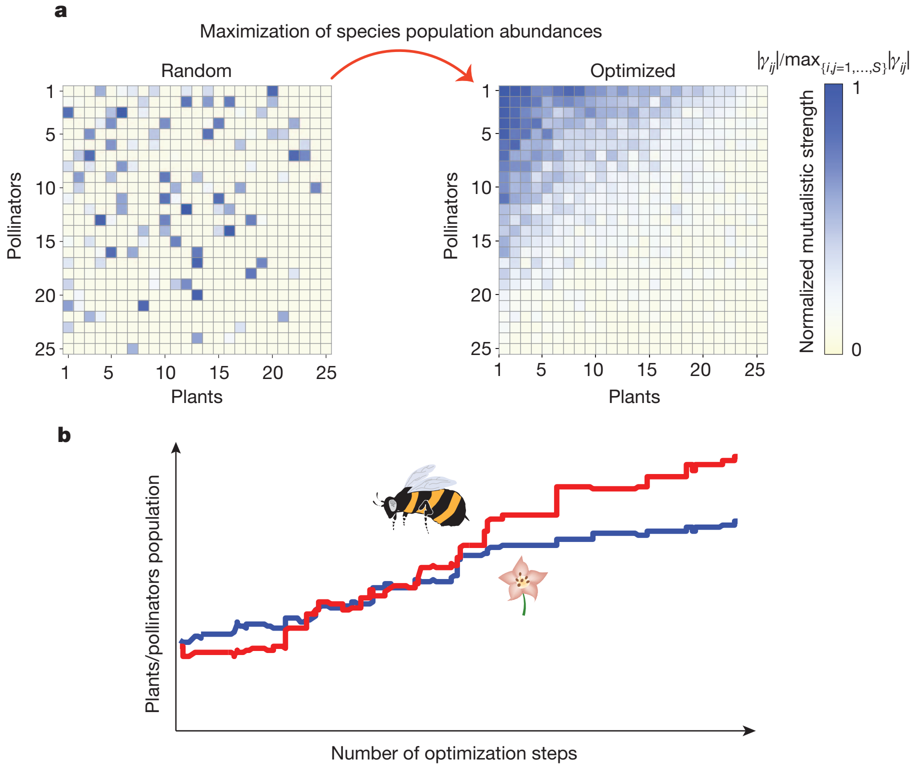


Figure 1 | The optimization principle. **a**, The maximization of species abundance leads to networks with a nested architecture. The optimal interaction matrix shown is the typical architecture resulting from averages performed over 100 optimal networks starting from random realizations. The blue scale is a measure of the average mutualistic strength normalized with respect to the maximum-strength interaction. **b**, Because of mutualism, the optimization of the abundances of the individual species involved in the interaction rewiring results in an overall increase of the total population of both pollinators (in red) and plants (in blue). The curves represent the result of a typical run (no average is involved). Simulations presented here are obtained with Holling type II dynamics and parameters $S = 50$, $C_\Omega = C_\Gamma = 4/S^{0.8}$ and $\sigma_\Omega = \sigma_\Gamma < \sigma_c$ (see Methods Summary).

nestedness distributions of the optimized mutualistic networks shift markedly to higher values than their random network counterparts (see Figs 2a, b and 3). Monte Carlo simulations substantiate the strong correlation between the total population and the nestedness of the mutualistic interaction network (Fig. 2c). Our analytic calculations show that, for identical increments in population abundances, a community characterized by weak mutualism has a larger increase in nestedness than one with strong mutualistic interactions (Supplementary Information). This result suggests that, when the mutualistic interactions are strong, the network architecture may have a less crucial role than in the regime of weak mutualistic interactions, wherein optimal tuning of the architecture could lead to considerable beneficial effects for the community.

Our results are very robust and do not depend on the details of the optimization algorithm, the initial condition or the transient dynamics. In addition to simulations of mutualistic communities starting with random interactions networks and then 'reorganizing' towards a more

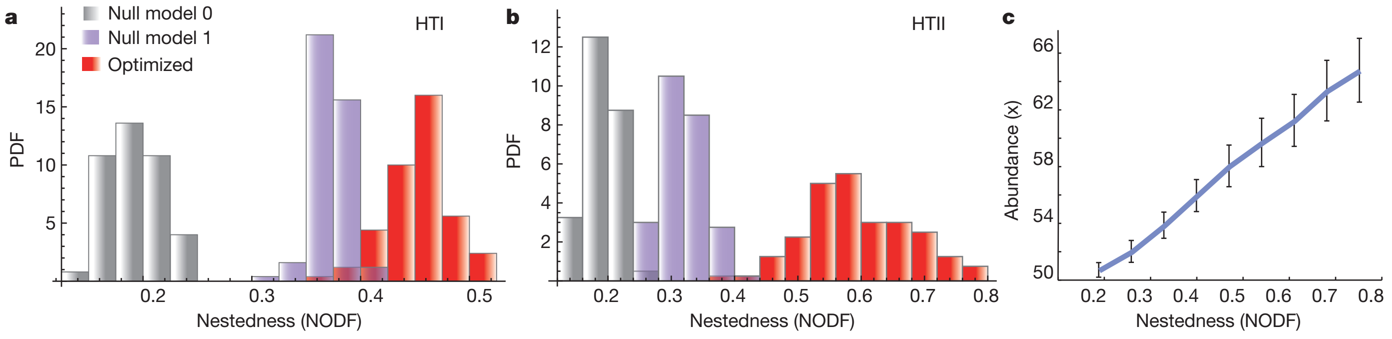


Figure 2 | Relationship between nestedness and species abundance. **a**, **b**, Histograms of the nestedness probability density for optimized mutualistic networks (shown in red) using the individual species optimization algorithm using Holling type I (HTI) (**a**) and Holling type II (HTII) (**b**) saturating functions. The histograms for the corresponding null model randomizations are also shown. In null model 0 (refs 6, 27), we preserve the dimensions and the connectivity of the optimized interaction network M with a random placement of the edges. In null model 1 (refs 6, 27), we also conserve the average number of connections for each plant and insect. The plots are obtained using 100 realizations of the optimization algorithm presented in the main text. In each realization, a new initial interaction matrix, M, is extracted with the same average $\mu = 0$, variance $\sigma_\Omega = \sigma_\Gamma < \sigma_c$, and connectance $C_\Omega = C_\Gamma$ (see Methods). **c**, We consider interaction matrices ($S = 50$, $C_\Omega = C_\Gamma$ and $\sigma_\Omega = \sigma_\Gamma < \sigma_c$) with different values of nestedness and we calculate the stationary population associated with each one of them: the nestedness and the total abundance of individuals in the community are strongly correlated. Error bars represent the ± 1 standard deviation over 100 realizations. The connectance has been chosen to vary with the number of species as $C_\Gamma = 4S^{-0.8}$, obtained as a best fit to empirical data (Supplementary Information). Similar results are obtained for different parameter values and implementations of the optimization algorithm. PDF, probability density function. NODF, nestedness metric based on overlap and decreasing fill.

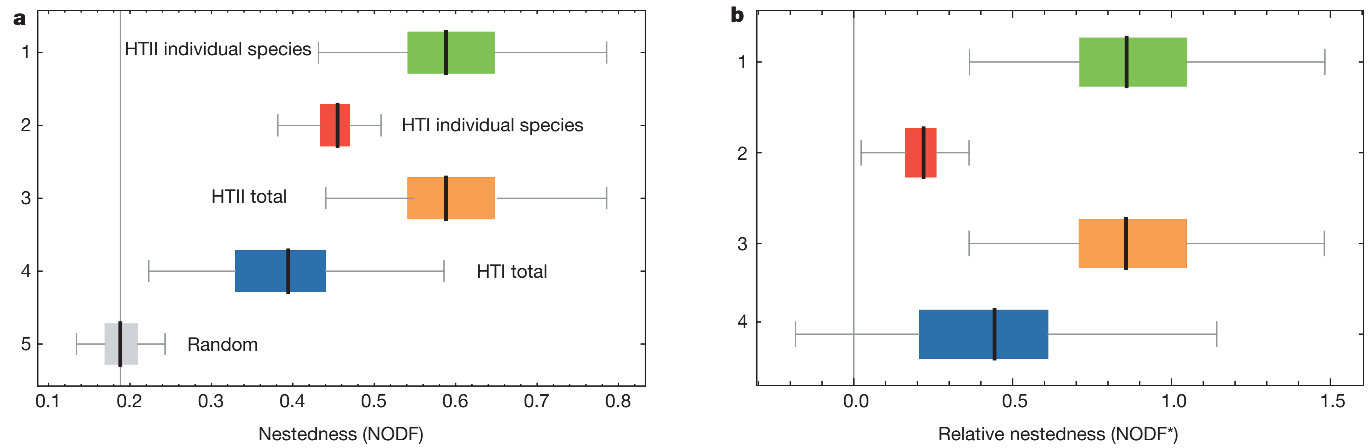


Figure 3 | Box–whisker plots of the degrees of nestedness for optimized networks. **a**, **b**, The ends of the whiskers represent the minimum and maximum, whereas the ends of the box are the first and third quartiles and the black bar denotes the median. The plots show the absolute nestedness (NODF) (**a**) and relative nestedness[10] (**b**) normalized to null model 1 (refs 6, 27) of 100 bipartite networks resulting from: species-level optimization (1, 2), community-level optimization (3, 4) and random mutualistic networks—null model 0 (refs 6, 27) (5). Parameters used here are $S = 50$, $C_\Omega = C_\Gamma = 4/S^{0.8}$ and $\sigma_\Omega = \sigma_\Gamma < \sigma_c$.

optimal state, we also implement a more realistic scenario in which mutualistic communities progressively assemble and are optimized over the course of evolutionary timescales[23,24]. Indeed, we find that the final result is the same, that is, the final optimized networks display a nested architecture (Supplementary Information). Notably, nested architectures in mutualistic communities could emerge from different initial conditions as a result of a rewiring of the interactions according to a variational principle aimed at maximizing either the fitness[25] of the individual insect/plant—whose surrogate is its species abundance—involved in the interaction swap (species-level optimization) or the fitness of the whole community, measured by the total population of all species (community-level optimization). The intriguing fact that these two optimization schemes lead to similar conclusions suggests that group selection mechanisms[26] may have had an important role in the evolution of cooperation among plant and pollinators[24].

Community persistence and stability are important dynamical properties characterizing ecological networks, but the way in which the two are related in real systems is far from trivial[21,22]. It has been suggested that mutualistic network structures lead to high community persistence[5,10]. Persistence, however, is only defined for systems out of their steady state[7,21], and is sensitive to initial conditions, transients dynamics, and to the system's distance from stationarity[11,13]. Here we focus on the study of community stability for the optimized stationary networks. Using perturbative expansion techniques, we analytically find that the abundance of the rarest species controls the stability of mutualistic communities (see Supplementary Information for mathematical details). Moreover, the optimized networks result in spontaneous symmetry breaking with more-abundant generalist species and less-abundant specialist species (Fig. 4a). The relatively low abundances of the specialist species make them more vulnerable to extinction and results in correspondingly lower community resilience, as measured by the maximum real part of the eigenvalues of the community matrix community (Fig. 4b). The advantage of having a high total population leading to increased robustness against extinction due to demographic fluctuations carries with it the cost associated with a lower resilience—the optimized network recovers from perturbations on a longer timescale than its random counterpart[12,14,27] (Fig. 4c).

Several ecological factors, as well as evolutionary history, contribute to shaping empirical networks. Here we have shown how binary nested network architecture could emerge as a consequence of an optimization process or variational principle. An interesting unexplored issue is an analysis of emergent quantitative nestedness[12], that is, the organization of the interaction intensities in the optimized networks along with a comparison to empirical data in mutualistic networks[12]. The

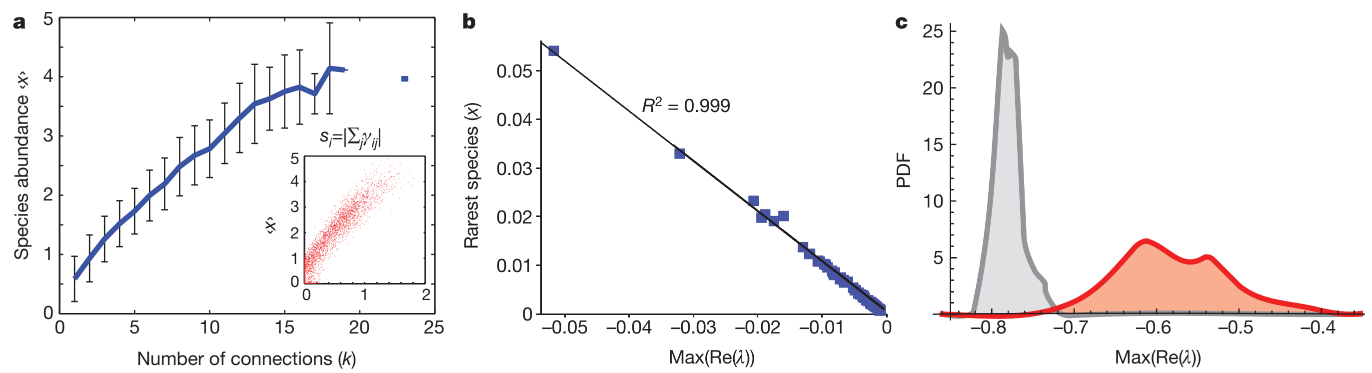


Figure 4 | Optimized networks are less resilient. **a**, Average species abundance $<x>$ as a function of the number of mutualistic partners of a species. The error bars represent the ±1 standard deviation confidence interval. Generalist species with more connections are, on average, more abundant than specialist species with fewer connections. The red points in the inset depict $<x>$ as a function of the mutualistic strength s. **b**, Relationship between the abundance of the rarest species and system resilience given by the largest among the real parts (closest to zero) of the eigenvalues of the linearized stability matrix. The grey line shows a linear fit ($R^2 = 0.999$). **c**, Probability density function of the largest among the real parts of the eigenvalues—Max(Re(λ))—of the optimized community stability matrix (red curve) and of the corresponding initial random networks (grey curve). Optimized networks are less resilient than their random counterparts. The plots are obtained from 100 realizations of the community-optimization algorithm performed with HTI saturating function $S = 50$, $C = 4/S^{0.8}$ and $\sigma_\Omega = \sigma_\Gamma < \sigma_c$.

framework proposed here ought to be applicable for investigating the possible driving forces sculpting mutualistic network architectures in a variety of systems ranging from social[28] to economic[29] and other biological[30] (for example, protein interaction) networks.

METHODS SUMMARY

Interaction networks. The initial interaction matrix M is composed of four blocks: two diagonal blocks describing direct competition among plants (Ω_{PP}) and pollinators (Ω_{AA}) and two off-diagonal blocks characterizing the mutualistic interactions between n_P plants and n_A pollinators (Γ_{PA}), and vice versa (Γ_{AP}). The total possible number of mutualistic interactions in each of these two latter blocks is equal to $n_A \times n_P$. The connectance C_Γ (or C_Ω) represents the fraction of the mutualistic (competitive) interactions that are non-zero. Mutualistic interaction intensities γ_{ij}^{AP} and γ_{ij}^{PA} represent the increase of the growth rate of animal (or plant) species i per unit of plant (animal) biomass j and they are assigned randomly from the distribution $|N(0,\sigma_\Gamma)|$, whereas the competitive interactions are distributed as $-|N(0,\sigma_\Omega)|$. Here, $N(\mu,\sigma)$ is the normal distribution with the mean μ and variance σ^2 chosen to have stability of the underlying population dynamics, that is, $\sigma < \sigma_c$, where σ_c is the critical strength threshold above which, with high probability, no stable fixed point dynamics exist[14].

Optimization algorithm. We start with an existing network; select a species, j, randomly and an existing link to one of its partner species k; we attempt a rewire between the j–k and the j–m links (where m is a potential alternative mutualistic partner), that is γ_{jk} is interchanged with γ_{jm}. If the j–m link already exists, that is, γ_{jm} is different from zero, the switch leads to an interchange of interaction strengths; otherwise the swap corresponds to rewiring the j–k link to j–m. The switch is accepted if and only if it does not lead to a decrease of the population abundance of species j in the steady state of the new network configuration. See Supplementary Information for details.

Received 15 January; accepted 9 July 2013.

1. Pascual, M. & Dunne, J. *Ecological Networks: Linking Structure to Dynamics in Food Webs* (Oxford Univ. Press, 2006).
2. Montoya, J. M., Pimm, S. L. & Sole, R. V. Ecological networks and their fragility. *Nature* **442,** 259–264 (2006).
3. Bascompte, J. & Jordano, P. Plant–animal mutualistic networks: the architecture of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **38,** 567–593 (2007).
4. Almeida-Neto, M., Guimara, P., Guimara, P. R. & Ulrich, W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117,** 1227–1239 (2008).
5. Bastolla, U. *et al.* The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* **458,** 1018–1020 (2009).
6. Bascompte, J., Jordano, P., Melian, C. J. & Olesen, J. M. The nested assembly of plant–animal mutualistic networks. *Proc. Natl Acad. Sci. USA* **100,** 9383–9387 (2003).
7. Saavedra, S., Stouffer, D. B., Uzzi, B. & Bascompte, J. Strong contributors to network persistence are the most vulnerable to extinction. *Nature* **478,** 233–235 (2011).
8. Krishna, A., Guimara, P. R., Jordano, P. & Bascompte, J. A neutral-niche theory of nestedness in mutualistic networks. *Oikos* **117,** 1609–1618 (2008).
9. Okuyama, T. & Holland, J. N. Network structural properties mediate the stability of mutualistic communities. *Ecol. Lett.* **11,** 208–216 (2008).
10. Thébault, E. & Fontaine, C. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* **329,** 853–856 (2010).
11. James, A., Pitchford, J. W. & Plank, M. J. Disentangling nestedness from models of ecological complexity. *Nature* **487,** 227–230 (2012).
12. Staniczenko, P. P., Kopp, J. C. & Allesina, S. The ghost of nestedness in ecological networks. *Nature Commun.* **4,** 1391 (2013).
13. Allesina, S. The more the merrier. *Nature* **487,** 175–176 (2012).
14. Allesina, S. & Tang, S. Stability criteria for complex ecosystems. *Nature* **483,** 205–208 (2012).
15. Campbell, C., Yang, S., Shea, K. & Albert, R. Topology of plant-pollinator networks that are vulnerable to collapse from species extinction. *Phys. Rev. E* **86,** 021924 (2012).
16. Thompson, A. R., Nisbet, R. M. & Schmitt, R. J. Dynamics of mutualist populations that are demographically open. *J. Anim. Ecol.* **75,** 1239–1251 (2006).
17. Holland, J. N., DeAngelis, D. L. & Bronstein, J. L. Population dynamics and mutualism: functional responses of benefits and costs. *Am. Nat.* **159,** 231–244 (2002).
18. Twardochleb, L. A. N. o. v. a. k. M. & Moore, J. W. Using the functional response of a consumer to predict biotic resistance to invasive prey. *Ecol. Appl.* **22,** 1162–1171 (2012).
19. Zhang, F., Hui, C. & Terblanche, J. S. An interaction switch predicts the nested architecture of mutualistic networks. *Ecol. Lett.* **14,** 797–803 (2011).
20. Douglas, M. W. Adaptation and habitat selection in the eco-evolutionary process. *Proc. R. Soc. B* **278,** 2401–2411 (2011).
21. McCann, K. S. The diversity–stability debate. *Nature* **405,** 228–233 (2000).
22. Ives, A. R. & Carpenter, S. R. Stability and diversity of ecosystems. *Science* **317,** 58–62 (2007).
23. Guimarães, P. R. *et al.* Interaction intimacy affects structure and coevolutionary dynamics in mutualistic networks. *Curr. Biol.* **17,** 1797–1803 (2007).
24. Guimarães, J., Jordano, P. & Thompson, J. N. Evolution and coevolution in mutualistic networks. *Ecol. Lett.* **14,** 877–885 (2011).
25. Nowak, M. A., Tarnita, C. E. & Wilson, E. O. The evolution of eusociality. *Nature* **466,** 1057–1062 (2010).
26. Nowak, M. A. Five rules for the evolution of cooperation. *Science* **314,** 1560–1563 (2006).
27. Ulrich, W., Almeida-Neto, W. & Gotelli, J. N. A consumer's guide to nestedness analysis. *Oikos* **118,** 3–17 (2009).
28. Saavedra, S., Reed-Tsochas, F. & Uzzi, B. A simple model of bipartite cooperation for ecological and organizational networks. *Nature* **457,** 463–466 (2009).
29. May, R. M., Levin, S. A. & Sugihara, G. Complex systems: ecology for bankers. *Nature* **451,** 893–895 (2008).
30. Proulx, S. R., Promislow, D. E. & Phillips, P. C. Network thinking in ecology and evolution. *Trends Ecol. Evol.* **20,** 345–353 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements A.M. and S.S. acknowledge the Cariparo foundation for financial support. We thank S. Allesina, T. Cooke, J. Grilli and L. Mari for discussions and Studio 7 a.m. for graphics support.

Author Contributions S.S. carried out the numerical calculations and the data analysis. All the authors contributed to other aspects of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.S. (suweis@pd.infn.it) or A.M. (amos.maritan@pd.infn.it).

OPEN
doi:10.1038/nature12326

Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*

Jean-François Flot[1,2,3,4,5,6], Boris Hespeels[1,2], Xiang Li[1,2], Benjamin Noel[3], Irina Arkhipova[7], Etienne G. J. Danchin[8,9,10], Andreas Hejnol[11], Bernard Henrissat[12], Romain Koszul[13], Jean-Marc Aury[3], Valérie Barbe[3], Roxane-Marie Barthélémy[14], Jens Bast[15], Georgii A. Bazykin[16,17], Olivier Chabrol[14], Arnaud Couloux[3], Martine Da Rocha[8,9,10], Corinne Da Silva[3], Eugene Gladyshev[7], Philippe Gouret[14], Oskar Hallatschek[6,18], Bette Hecox-Lea[7,19], Karine Labadie[3], Benjamin Lejeune[1,2], Oliver Piskurek[20], Julie Poulain[3], Fernando Rodriguez[7], Joseph F. Ryan[11], Olga A. Vakhrusheva[16,17], Eric Wajnberg[8,9,10], Bénédicte Wirth[14], Irina Yushenova[7], Manolis Kellis[21], Alexey S. Kondrashov[16,22], David B. Mark Welch[7], Pierre Pontarotti[14], Jean Weissenbach[3,4,5], Patrick Wincker[3,4,5], Olivier Jaillon[3,4,5,21]* & Karine Van Doninck[1,2]*

Loss of sexual reproduction is considered an evolutionary dead end for metazoans, but bdelloid rotifers challenge this view as they appear to have persisted asexually for millions of years[1]. Neither male sex organs nor meiosis have ever been observed in these microscopic animals: oocytes are formed through mitotic divisions, with no reduction of chromosome number and no indication of chromosome pairing[2]. However, current evidence does not exclude that they may engage in sex on rare, cryptic occasions. Here we report the genome of a bdelloid rotifer, *Adineta vaga* (Davis, 1873)[3], and show that its structure is incompatible with conventional meiosis. At gene scale, the genome of *A. vaga* is tetraploid and comprises both anciently duplicated segments and less divergent allelic regions. However, in contrast to sexual species, the allelic regions are rearranged and sometimes even found on the same chromosome. Such structure does not allow meiotic pairing; instead, we find abundant evidence of gene conversion, which may limit the accumulation of deleterious mutations in the absence of meiosis. Gene families involved in resistance to oxidation, carbohydrate metabolism and defence against transposons are significantly expanded, which may explain why transposable elements cover only 3% of the assembled sequence. Furthermore, 8% of the genes are likely to be of non-metazoan origin and were probably acquired horizontally. This apparent convergence between bdelloids and prokaryotes sheds new light on the evolutionary significance of sex.

With more than 460 described species[4], bdelloid rotifers (Fig. 1) represent the highest metazoan taxonomic rank in which males, hermaphrodites and meiosis are unknown. Such persistence and diversification of an ameiotic clade of animals are in contradiction with the supposed long-term disadvantages of asexuality, making bdelloids an 'evolutionary scandal'[5]. Another unusual feature of bdelloid rotifers is their extreme resistance to desiccation at any stage of their life cycle[6], enabling these microscopic animals to dwell in ephemeral freshwater habitats such as mosses, lichens and forest litter; this ability is presumably the source of their extreme resistance to ionizing radiation[7].

We assembled the genome of a clonal *A. vaga* lineage into separate haplotypes with a N_{50} of 260 kilobases (kb) (that is, half of the assembly was composed of fragments longer than 260 kb). Assembly size was 218 megabases (Mb) but 26 Mb of the sequence had twice the average sequencing coverage, suggesting that some nearly identical regions were not resolved during assembly (Supplementary Fig. 3); hence, the total genome size is likely to be 244 Mb, which corresponds to the estimate obtained independently using fluorometry (Supplementary Note C2). Annotation of the complete assembly (including all haplotypes) yielded 49,300 genes. Intragenomic sequence comparisons revealed numerous homologous blocks with conserved gene order (colinear regions). For each such block we computed the per-site synonymous divergence (K_s) and a colinearity metric defined as the fraction of colinear genes. Colinear blocks fell into two groups (Fig. 2a): a group characterized by high colinearity and low average synonymous divergence, and a group characterized by lower colinearity and higher synonymous divergence. The presence of two classes of colinear blocks is consistent with a tetraploid structure comprised of alleles (recent homologues) and ohnologues (ancient homologues formed by genome duplication). Allelic pairs of coding sequences are on average 96.2%

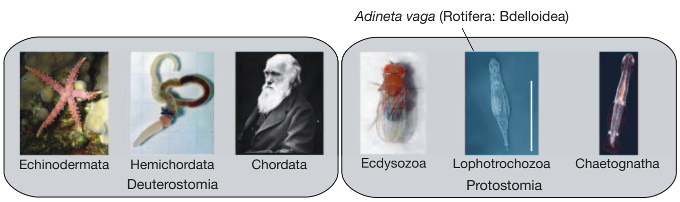


Figure 1 | Position of bdelloid rotifers among metazoans. Bdelloid rotifers ('leech-like wheel-bearers') are a clade of microscopic animals (scale bar, 100 μm) within the phylum Rotifera. Photographs of Hemichordata (*Saccoglossus*), Chordata (*Homo*) and Ecdysozoa (*Drosophila*) courtesy of David Remsen (MBL), John van Wyhe (http://darwin-online.org.uk) and André Karwath, respectively.

[1]University of Namur, Department of Biology, URBE, Laboratory of Evolutionary Genetics and Ecology, 5000 Namur, Belgium. [2]Namur Research Institute for Life Sciences (NARILIS), 5000 Namur, Belgium. [3]CEA-Institut de Génomique, GENOSCOPE, Centre National de Séquençage, 2 rue Gaston Crémieux, CP5706, 91057 Evry Cedex, France. [4]Université d'Evry, UMR 8030, CP5706, 91057 Evry Cedex, France. [5]Centre National de la Recherche Scientifique (CNRS), UMR 8030, CP5706, 91057 Evry Cedex, France. [6]Max Planck Institute for Dynamics and Self-Organization, Biological Physics and Evolutionary Dynamics, Bunsenstraße 10, 37073 Göttingen, Germany. [7]Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA. [8]INRA, UMR 1355 ISA, Institut Sophia Agrobiotech, 400 route des Chappes, 06903 Sophia-Antipolis, France. [9]CNRS, UMR 7254 ISA, Institut Sophia Agrobiotech, 400 route des Chappes, 06903 Sophia-Antipolis, France. [10]Université de Nice Sophia-Antipolis, UMR ISA, Institut Sophia Agrobiotech, 400 route des Chappes, 06903, Sophia-Antipolis, France. [11]Sars International Centre for Marine Molecular Biology, University of Bergen, 5008 Bergen, Norway. [12]Architecture et Fonction des Macromolécules Biologiques, Aix-Marseille University, CNRS UMR 7257, 13288 Marseille, France. [13]Groupe Spatial regulation of genomes, CNRS UMR 3525, Department of Genomes and Genetics, Institut Pasteur, 75724 Paris, France. [14]LATP UMR-CNRS 7353, Evolution Biologique et Modélisation, Aix-Marseille University, 13331 Marseille cedex 3, France. [15]J.F. Blumenbach Institute of Zoology and Anthropology, University of Göttingen, 37073 Göttingen, Germany. [16]Department of Bioengineering and Bioinformatics, M.V. Lomonosov Moscow State University, Leninskye Gory 1-73, Moscow, 119991, Russia. [17]Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Bolshoi Karetny pereulok 19, Moscow, 127994, Russia. [18]Department of Physics, University of California, Berkeley, California 94720, USA. [19]Department of Biology, Northeastern University, Boston, Massachusetts 02115, USA. [20]Courant Research Centre Geobiology, Georg-August-Universität Göttingen, Goldschmidtstraße 3, Göttingen 37077, Germany. [21]MIT Computer Science and Artificial Intelligence Laboratory, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02139, USA. [22]Life Sciences Institute and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109-2216, USA.
*These authors contributed equally to this work.

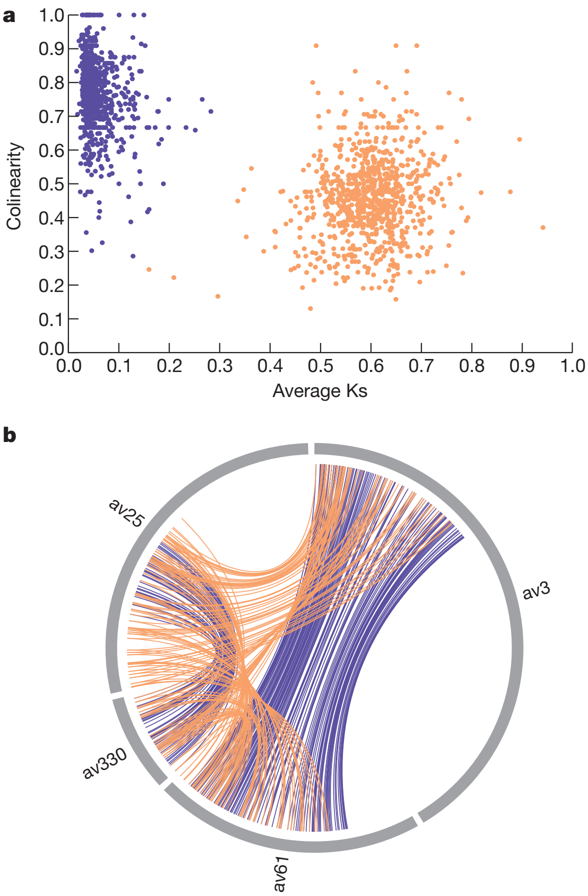


Figure 2 | A locally tetraploid genome. a, Analysis of intragenomic synteny reveals two groups of colinear regions: alleles (in violet, regions characterized by a high fraction of colinear genes and low average Ks, that is, synonymous divergence) and ohnologues (in orange, with lower colinearity but higher Ks). **b**, Example of a genomic quartet of four scaffolds: allelic gene pairs are connected with violet curves and ohnologous gene pairs with orange curves.

identical at the nucleotide level (median = 98.6%) versus 73.6% (median = 75.1%) for ohnologous pairs. Nearly 40% (84.5 Mb) of the assembled genome sequence is organized in quartets of four homologous regions A_1, A_2, B_1 and B_2, of which A_1–A_2 and B_1–B_2 are two pairs of alleles and As are ohnologous to Bs[8] (Fig. 2b).

We found evidence of genomic palindromes up to 705 kb in length and involving up to 148 genes. The *A. vaga* genome contains at least 17 such palindromic regions (Fig. 3a) reminiscent of those reported in the Y chromosomes of primates[9]. In all 17 cases, the arms of the palindromes present the colinearity and divergence signatures of allelic regions and do not have other allelic duplicates in the assembly, suggesting that they arose by inter-allelic rearrangements rather than by local duplications. In addition to these 17 inverted repeats, we observed three direct repeats that present the signatures of allelic blocks and involve up to 50 genes (Fig. 3a). The cumulative length of the assembly fragments (scaffolds) bearing these 20 allelic rearrangements is 7.5 Mb or 3.5% of the genome sequence. Allelic regions that are found on the same chromosome clearly cannot segregate during meiosis. Moreover, we found hundreds of colinearity breakpoints between allelic regions, and the total length of the scaffolds that have no full-length homologue in the assembly due to these breakpoints exceeds 109 Mb or 51% of the genome assembly (including 91 of the 100 largest scaffolds, Fig. 3b and Supplementary Fig. 10). As a result, it is impossible to split the assembled genome of *A. vaga* into haploid sets: the apparent ploidy level of *A. vaga* is scale-dependent, with a tetraploid structure at gene scale versus chromosome-scale haploidy. Such relaxation of constraints on genome structure is reminiscent of other mitotic lineages such as cancer cells[10] and somatic tissues[11].

It has been proposed that, in the absence of meiosis, alleles accumulate mutations independently from one another, to the point that ancient asexuals may harbour genome-wide allele sequence divergence (ASD)[12] larger than inter-individual differences (the so-called 'Meselson effect'). However, the average inter-allelic divergence of *A. vaga* is only 4.4% at the nucleotide level (3% when looking at synonymous divergence), which falls in the upper range reported for sexually reproducing species[13]. The absence of genome-wide ASD could be explained by low mutation rates and/or by frequent mitotic recombination (such as gene conversion resulting from DNA repair)[12]. Although there is no evidence of reduced mutation rates in bdelloid rotifers compared with their cyclically sexual sister clade the monogononts[14], we found strong signatures

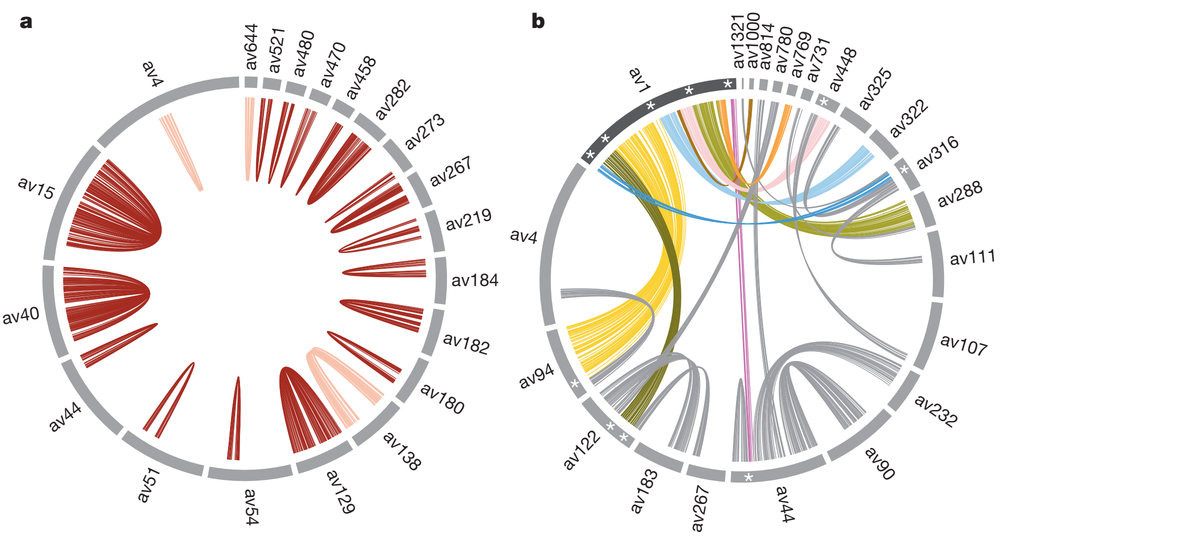


Figure 3 | A genome structure incompatible with conventional meiosis. a, In twenty cases, allelic regions are found to occur on the same chromosome. All curves shown connect allelic gene pairs. On three scaffolds both allelic regions have the same orientation (direct repeats, in pink), whereas on the seventeen other scaffolds they are inverted (palindromes, in red). **b**, Local colinearity between alleles does not extend to chromosome scale. Colours are arbitrary and only allelic gene pairs are represented. Asterisks highlight colinearity breakpoints between scaffold av1 and its allelic partners av44, av94, av122, av316 and av448. Further examples for other scaffolds are shown on Supplementary Fig. 10.

of recent gene conversion events in the distribution of identity track lengths, that is, distances between consecutive mismatches (Fig. 4a and Supplementary Note E1). We calculated that the probability that a given base in the genome experiences gene conversion is at least one order of magnitude greater than its probability to mutate (Supplementary Note E1), suggesting that homologous regions in the genome of *A. vaga* undergo concerted evolution[15]. Homogenization through gene conversion may either expose new mutations to selection by making them homozygous or remove them as they get overwritten with the other allelic version (Fig. 4b), thereby slowing Muller's ratchet (that is, the irreversible accumulation of detrimental mutations in asexual populations of finite sizes, Supplementary Note E2 and Supplementary Fig. 11).

Over 8% of the genes of *A. vaga* are much more similar to non-metazoan sequences in GenBank than to metazoan ones (AI log score > 45 (ref. 16), Supplementary Note E4) and were therefore probably acquired through horizontal gene transfer (HGT). This class of genes has significantly fewer introns per kilobase of coding sequence compared with probable core metazoan genes (AI ≤ −45, Supplementary Table 2). More than 20% of genes with AI > 45 are found in quartets (groups of four homologous copies in conserved syntenic regions) and were therefore probably incorporated into the rotifer genome before the establishment of tetraploidy, which itself pre-dates the divergence of extant bdelloid families[8]. The higher the number of copies of a putative HGT gene, the higher its number of introns and the closer its guanine–cytosine (GC) content to the *A. vaga* genome average (Supplementary Fig. 22), which suggests that these parameters reflect the age of acquisition. We also noticed signatures of possibly very recent HGTs: 60 genes with AI > 45 are present in only one copy (with normal coverage), have no intron and have a GC content that is more than 1% above or below the genome average (the same scaffolds also bear genes of probable metazoan origin with AI < 0). In summary, there seems to be an ancient but still ongoing process of HGT at a level comparable to some bacteria[17].

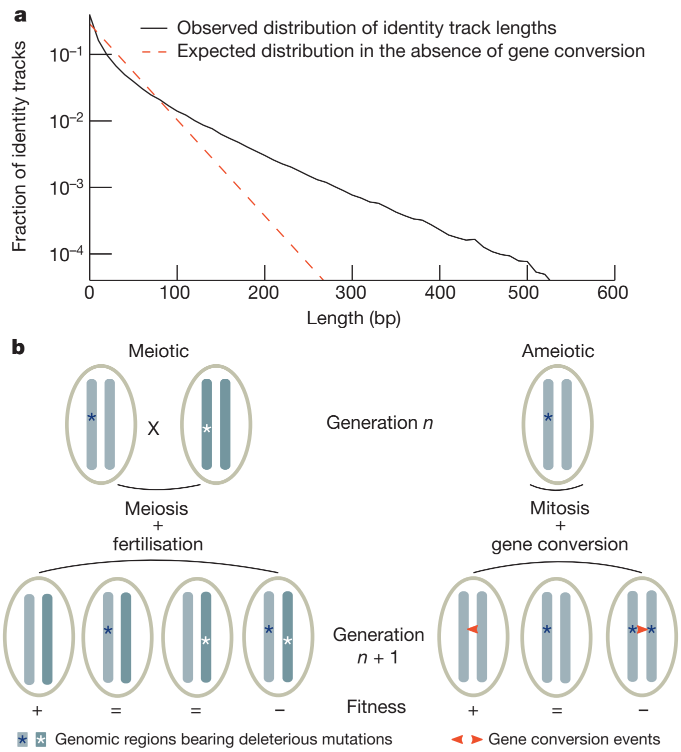


Figure 4 | Gene conversion and its evolutionary consequences in ameiotic organisms. **a**, Evidence for gene conversion between allelic regions. If we suppose that mutations happen at random in a Poisson process of parameter $1/M$ (where M is the average distance between mutations), then the distance between two consecutive mismatches follows a negative exponential distribution where the proportion of identity tracks of length x equals $e^{-x/M}/M$. Comparison of the observed distribution of identity track lengths with this theoretical distribution reveals a deficit of short tracks and an excess of long tracks, as expected in case of gene conversion. The same pattern was observed when gene-coding regions were excluded from the analysis (data not shown), thereby ruling out a confounding effect of selection. **b**, In sexual organisms, meiotic recombination can generate offspring with fewer or more deleterious mutations (hence increasing or decreasing fitness) than the previous generation. The same outcome is expected in ameiotic organisms that experience gene conversion: a deleterious allele may be overwritten by a beneficial or neutral one, resulting in an increase in fitness, or may overwrite it, resulting in decreased fitness.

Some theories predict that transposable elements should be either absent from the genomes of asexuals[18] or undergo unrestrained expansion after the switch to asexuality, potentially leading to species extinction unless transposable element proliferation is prevented[19]. We found that transposable elements cover about 3% of the *A. vaga* genome, which is less than the percentage reported in most other metazoans (including the genome of the obligate parthenogenetic nematode *Meloidogyne incognita*, 36% of which is made up of repetitive elements[20]). Another surprising feature is the high diversity of transposable-element families and the extremely low copy numbers observed for each of them (Supplementary Table 3). Out of 255 families, the overwhelming majority (209) are represented by only one or two full-length copies (for 24 families, no full-length copies could be identified), and for each full-length copy there are, on average, only about ten times as many transposable-element fragments. This relatively low abundance of decayed copies and the fact that long-terminal-repeat (LTR) retrotransposons have identical or nearly identical LTRs (Supplementary Table 4) suggest that most low-copy-number families represent recent arrivals. This is consistent with an ongoing process of acquisition of transposable elements by HGT.

This hypothesis is further supported by the significantly higher density of transposable elements observed around HGTs and vice-versa (Supplementary Note E5). If *A. vaga* has been acquiring transposable elements by HGT, a question that arises is what keeps their number lower than in most other metazoans. Many fragmented copies have apparently been formed through microhomology-mediated deletions. Excision of LTR retrotransposons has also been occurring through LTR–LTR recombination, leaving behind numerous solo LTRs: for example, two *Juno1* insertions, *Juno1.1* and *Juno1.2*, which were present as full-length copies in the 2006 *A. vaga* fosmid library[21], exist in the current assembly only as solo LTRs (in the same genomic environments and with the same target site duplications). Finally, there is evidence for expansion and diversification of the RNA-mediated silencing machinery. In addition to Dicer1 proteins, which are shared by all metazoans, *A. vaga* possesses a deep-branching Dicer-like clade with uncertain taxonomic placement (Supplementary Fig. 20). The Argonaute/Piwi and RNA-directed RNA polymerase (RdRP) families are also expanded (Supplementary Figs 18 and 19). It is plausible that these proteins participate in epigenetic silencing of transposable elements (as was recently observed for single-copy transgenes in *Caenorhabditis elegans*[22]), thereby preventing horizontally transferred transposable elements from multiplying upon arrival.

Overall, the genome of *A. vaga* comprises more genes than usually reported for metazoans (Supplementary Note F2), as its haplotypes were assembled separately. Even taking this into account, the gene repertoire of *A. vaga* features expansion of several gene families. For example, the genome of *A. vaga* comprises 284 homeobox superclass genes, mostly found in four copies (quartets) but not organized in clusters; very few ohnologues have been lost, resulting in more homeobox genes than in any other metazoan genome sequenced (Supplementary Note F5). Genes putatively related to oxido-reduction processes are substantially more abundant in *A. vaga* than in other metazoan species, and most of the corresponding genes appear to be constitutively expressed (Supplementary Table 9). This is consistent with the recent report of an effective antioxidant protection system

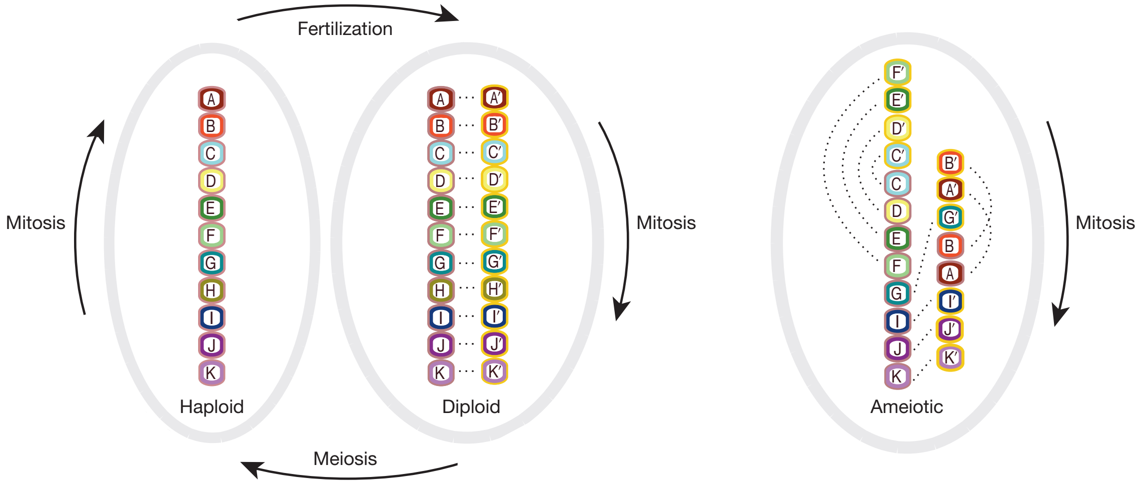


Figure 5 | Meiotic versus ameiotic genome structures. Genes are represented with letters, and dashed lines connect allelic gene pairs. A meiotic genome (left) alternates between a haploid phase (in which a single allele of each gene is present) and a diploid phase (in which the genes are present in two allelic versions arranged colinearly on homologous chromosomes). In the ameiotic genome of *A. vaga* (right), alleles are distributed in blocks that are shuffled across chromosomes, resulting notably in intrachromosomal repeats (direct or inverted). As a consequence, chromosomes have no homologues and cannot be paired.

in bdelloid rotifers[23]. Carbohydrate-active enzymes (CAZymes) in the genome of *A. vaga* are also notably diverse and abundant, with 1,075 genes falling into 202 characterized families. With 623 glycoside hydrolases (involved in the hydrolysis of sugar bonds) and 412 glycosyltransferases (responsible for building sugar bonds), the CAZyme richness of *A. vaga* ranks highest among metazoans and is only comparable to some plants such as poplars[24]. *A. vaga* has the richest repertoire of glycoside hydrolases of any organism sequenced so far, hinting at a diversity of feeding habits; 52% of the CAZymes have an AI > 45 and were therefore probably acquired through horizontal gene transfer.

A. vaga has lost 1,250 genes compared with the inferred last common ancestor of Protostomia, the genome of which comprised at least 7,844 unique protein-coding genes (Supplementary Note E6). A total of 137 PFAM domains typically present in metazoans could not be detected in the assembled genome sequence (Supplementary Data 10). Of particular interest are missing domains involved in reproductive processes (Supplementary Note F1); for example, the *Zona pellucida*-like domain (notably found in sperm-binding proteins[25]) is present in an average of 36 copies in metazoan genomes but is absent in *A. vaga*. In contrast, we found multiple copies of most metazoan genes involved in DNA repair and homologous recombination, including a considerably divergent *Spo11* but no *Rad52* and *Msh3*.

To conclude, our analysis of a lineage of the bdelloid rotifer *Adineta vaga* reveals positive evidence for asexual evolution: its genome structure does not allow pairing of homologous chromosomes and therefore seems incompatible with conventional meiosis (Fig. 5). However, we cannot rule out that other forms of recombination occur in bdelloid populations in ways that do not require homologous pairing, such as parasexuality[26]. The high number of horizontally acquired genes, including some seemingly recent ones, suggests that HGTs may also be occurring from rotifer to rotifer. It is plausible that the repeated cycles of desiccation and rehydration experienced by *A. vaga* in its natural habitats have had a major role in shaping its genome: desiccation presumably causes DNA double-strand breaks, and these breaks that allow integration of horizontally transferred genetic material also promote gene conversion when they are repaired. Hence, the homogenizing and diversifying roles of sex may have been replaced in bdelloids by gene conversion and horizontal gene transfer, in an unexpected convergence of evolutionary strategy with prokaryotes.

METHODS SUMMARY

Genomic DNA was extracted from laboratory cultures of a clonal *A. vaga* lineage and shotgun-sequenced using 454 and Illumina platforms at respective coverage of 25 and 440 times (using both single reads and mate reads from inserts up to 20 kb). The 454 reads were assembled into contigs using MIRA[27]; the contigs obtained were corrected using single Illumina reads and linked into scaffolds using paired Illumina reads[28] (Supplementary Table 1). We annotated protein-coding genes by integrating evidence from RNA sequencing, *ab initio* predictions and comparison with UniProt. Most synteny and Ka/Ks (non-synonymous divergence/synonymous divergence) analyses were performed using the package MCScanX[29] and synteny plots were drawn using Circos[30].

Received 21 November 2012; accepted 30 May 2013.

Published online 21 July; corrected online 21 August 2013 (see full-text HTML version for details).

1. Danchin, E. G. J., Flot, J.-F., Perfus-Barbeoch, L. & Van Doninck, K. In *Evolutionary Biology — Concepts, Biodiversity, Macroevolution and Genome Evolution* (ed. Pontarotti, P.) 223–242 (Springer, 2011).
2. Hsu, W. S. Oogenesis in the Bdelloidea rotifer *Philodina roseola* Ehrenberg. *Cellule* **57,** 283–296 (1956).
3. Davis, H. A new *Callidina*: with the result of experiments on the desiccation of rotifers. *Month. Microscopical J.* **9,** 201–209 (1873).
4. Segers, H. Annotated checklist of the rotifers (Phylum Rotifera), with notes on nomenclature, taxonomy and distribution. *Zootaxa* **1564,** 1–104 (2007).
5. Maynard Smith, J. Contemplating life without sex. *Nature* **324,** 300–301 (1986).
6. Ricci, C. Anhydrobiotic capabilities of bdelloid rotifers. *Hydrobiologia* **387–388,** 321–326 (1998).
7. Gladyshev, E. & Meselson, M. Extreme resistance of bdelloid rotifers to ionizing radiation. *Proc. Natl Acad. Sci. USA* **105,** 5139–5144 (2008).
8. Hur, J. H., Van Doninck, K., Mandigo, M. L. & Meselson, M. Degenerate tetraploidy was established before bdelloid rotifer families diverged. *Mol. Biol. Evol.* **26,** 375–383 (2009).
9. Rozen, S. *et al.* Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423,** 873–876 (2003).
10. Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144,** 27–40 (2011).
11. Vijg, J. & Dollé, M. E. T. Large genome rearrangements as a primary cause of aging. *Mech. Ageing Dev.* **123,** 907–915 (2002).
12. Birky, C. W. Jr. Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics* **144,** 427–437 (1996).
13. Leffler, E. M. *et al.* Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol.* **10,** e1001388 (2012).
14. Welch, D. B. M. & Meselson, M. S. Rates of nucleotide substitution in sexual and anciently asexual rotifers. *Proc. Natl Acad. Sci. USA* **98,** 6720–6724 (2001).
15. Teshima, K. M. & Innan, H. The effect of gene conversion on the divergence between duplicated genes. *Genetics* **166,** 1553–1560 (2004).
16. Gladyshev, E. A., Meselson, M. & Arkhipova, I. R. Massive horizontal gene transfer in bdelloid rotifers. *Science* **320,** 1210–1213 (2008).
17. Syvanen, M. Evolutionary implications of horizontal gene transfer. *Annu. Rev. Genet.* **46,** 341–358 (2012).
18. Hickey, D. A. Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics* **101,** 519–531 (1982).
19. Arkhipova, I. & Meselson, M. Deleterious transposable elements and the extinction of asexuals. *Bioessays* **27,** 76–85 (2005).
20. Abad, P. *et al.* Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnol.* **26,** 909–915 (2008).

21. Gladyshev, E. A., Meselson, M. & Arkhipova, I. R. A deep-branching clade of retrovirus-like retrotransposons in bdelloid rotifers. *Gene* **390,** 136–145 (2007).
22. Shirayama, M. *et al.* piRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. *Cell* **150,** 65–77 (2012).
23. Krisko, A., Leroy, M., Radman, M. & Meselson, M. Extreme anti-oxidant protection against ionizing radiation in bdelloid rotifers. *Proc. Natl Acad. Sci. USA* **109,** 2354–2357 (2012).
24. Geisler-Lee, J. *et al.* Poplar carbohydrate-active enzymes. Gene identification and expression analyses. *Plant Physiol.* **140,** 946–962 (2006).
25. Bork, P. & Sander, C. A large domain common to sperm receptors (Zp2 and Zp3) and TGF-β type III receptor. *FEBS Lett.* **300,** 237–240 (1992).
26. Forche, A. *et al.* The parasexual cycle in *Candida albicans* provides an alternative pathway to meiosis for the formation of recombinant strains. *PLoS Biol.* **6,** e110 (2008).
27. Chevreux, B., Wetter, T. & Suhai, S. Genome sequence assembly using trace signals and additional sequence information. *Proc. German Conf. Bioinf.* **99,** 45–56 (1999).
28. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27,** 578–579 (2011).
29. Wang, Y. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40,** e49 (2012).
30. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19,** 1639–1645 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors would like to thank M. Meselson for his support during the initiation phase of this project and for inspiring us with his seminal works on bdelloid genetics. The authors are also grateful to M. Radman for useful discussions, M. Knapen and N. Debortoli for participating in laboratory work, M. Lliros for helping with Fig. 1, S. Henrissat for participating in CAZyme analyses, and S. Oztas, B. Vacherie, P. Lenoble and S. Mangenot for performing PCR validations of the assembly. This work was supported by Genoscope-CES (where most of the sequencing was performed), by US National Science Foundation grants MCB-0821956 and MCB-1121334 to I.A., by German Research Foundation grant HA 5163/2-1 to O.H., by grant 11.G34.31.0008 from the Ministry of Education and Science of the Russian Federation to A.S.K., by grant NSF CAREER number 0644282 to M.K., by US National Science Foundation grant MCB-0923676 to D.B.M.W., by FRFC grant 2.4.655.09.F from the Belgian Fonds National de la Recherche Scientifique (FNRS) and a start-up grant from the University of Namur to K.V.D.; J.F.F. and K.V.D. thank also J.-P. Descy (University of Namur) for funding support.

Author Contributions Bo.H., X.L., and B.N. are joint second authors; O.J. and K.V.D. are joint last authors. Bo.H., X.L., F.R. and B.H.L. maintained the rotifer cultures; Bo.H., X.L., F.R. and B.H.L. prepared the genomic DNA; X.L., D.B.M.W. and B.H.L. carried out gene expression experiments; Bo.H., X.L. and B.H.L. prepared complementary DNAs; K.L., J.P. and B.H.L. carried out the sequencing; J.F.F., A.C., V.B., O.J., B.N., J.M.A. and C.D.S. assembled the genome, validated the assembly and built the gene set; J.F.F., J.M.A., V.B., G.A.B., M.D.R., E.G.J.D., O.A.V., M.K., P.W., O.J. and K.V.D. analysed the genome structure; Bo.H., E.G.J.D., M.D.R., J.F.F., A.H., Be.H., B.H.L., R.K., B.L., J.F.R., F.R., A.S.K., E.W., D.B.M.W. and K.V.D. analysed the gene families; I.A., J.B., O.P. and I.Y. annotated and analysed the transposable elements; O.C., P.G., B.W., R.B., P.P. and K.V.D. carried out orthology analysis; I.A., E.G., E.G.J.D., P.G., B.W., F.R., D.B.M.W., P.P., J.F.F. and O.J. analysed the horizontal gene transfers; O.A.V., J.F.F., G.A.B., A.S.K. and D.B.M.W. analysed the signatures of gene conversion; O.H. modelled the effect of gene conversion on Muller's ratchet; J.F.F., O.J. and K.V.D. wrote the core of the manuscript, with contributions from I.A., E.G.J.D., A.H., B.N., O.H., Be.H., Bo.H., R.K., J.M.A., J.F.R., O.A.V., M.K., A.S.K., D.B.M.W., P.P. and P.W.; and P.W., J.W., R.B., D.B.M.W., P.P., O.J. and K.V.D. designed the project and acquired funding.

Author Information The sequencing reads and assembly are available at the Sequence Read Archive (accessions ERP002115 and SRP020364 for DNA, ERP002474 and SRP020358 for cDNA) and at the European Nucleotide Archive (accession CAWI000000000), respectively. The assembly and annotation can be browsed and downloaded at http://www.genoscope.cns.fr/adineta, whereas the result of the orthology analysis is accessible at http://ioda.univ-provence.fr/. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.J. (ojaillon@genoscope.cns.fr or ojaillon@mit.edu), J.F.F. (jean-francois.flot@ds.mpg.de) or K.V.D. (karine.vandoninck@fundp.ac.be).

LETTER

doi:10.1038/nature12330

Oxytocin enhances hippocampal spike transmission by modulating fast-spiking interneurons

Scott F. Owen[1], Sebnem N. Tuncdemir[2], Patrick L. Bader[1,2], Natasha N. Tirko[2], Gord Fishell[2] & Richard W. Tsien[1,2]

Neuromodulatory control by oxytocin is essential to a wide range of social[1,2], parental[3] and stress-related behaviours[4]. Autism spectrum disorders (ASD) are associated with deficiencies in oxytocin levels[5] and with genetic alterations of the oxytocin receptor (*OXTR*)[6]. Thirty years ago, Mühlethaler *et al.*[7] found that oxytocin increases the firing of inhibitory hippocampal neurons, but it remains unclear how elevated inhibition could account for the ability of oxytocin to improve information processing in the brain. Here we describe in mammalian hippocampus a simple yet powerful mechanism by which oxytocin enhances cortical information transfer while simultaneously lowering background activity, thus greatly improving the signal-to-noise ratio. Increased fast-spiking interneuron activity not only suppresses spontaneous pyramidal cell firing, but also enhances the fidelity of spike transmission and sharpens spike timing. Use-dependent depression at the fast-spiking interneuron–pyramidal cell synapse is both necessary and sufficient for the enhanced spike throughput. We show the generality of this novel circuit mechanism by activation of fast-spiking interneurons with cholecystokinin or channelrhodopsin-2. This provides insight into how a diffusely delivered neuromodulator can improve the performance of neural circuitry that requires synapse specificity and millisecond precision.

The CA1 region of hippocampus receives potent excitatory input from neighbouring area CA3 through the Schaffer Collateral (SC) pathway. Activation of SC axons evokes a monosynaptic excitatory postsynaptic potential (EPSP) onto CA1 pyramidal cells, as well as exciting a variety of CA1 interneurons. These interneurons then deliver a millisecond-delayed inhibitory postsynaptic potential (IPSP), termed feed-forward inhibition. Thus, both the stimulation threshold and the timing of spikes evoked in CA1 pyramidal cells by SC activation are dictated by a finely tuned balance of monosynaptic excitatory and disynaptic inhibitory inputs[8,9].

In agreement with previous results[8], we found that stimulation of the SC pathway in acute rat hippocampal slices evoked spikes with a short latency and moderate jitter (Fig. 1a). Strikingly, bath application of TGOT (Thr[4],Gly[7]-oxytocin, 200 nM), a specific agonist for oxytocin receptors, dramatically increased the probability of evoking a spike in the postsynaptic neuron from 0.50 to 0.82, while simultaneously suppressing the spontaneous activity of CA1 pyramidal cells by 57% from 1.4 Hz to 0.6 Hz (Fig. 1a–d). The combination of increased evoked spike probability (signal) and reduced spontaneous activity (noise) resulted in an enhanced signal-to-noise ratio. TGOT also reduced the latency and increased the temporal precision of evoked spikes (Fig. 1e, f).

TGOT increased the rate and amplitude of spontaneous inhibitory postsynaptic currents (IPSCs) onto CA1 pyramidal cells, as previously described[7,10] (Fig. 1g, Supplementary Fig. 1). Blockade by 10 μM bicuculline or by 100 nM tetrodotoxin indicated that these events were mediated by $GABA_A$ receptors and probably required an increase in interneuron firing rather than a change in spontaneous presynaptic release. The specific oxytocin receptor antagonist OTA (($d(CH_2)_5{}^1$,Tyr(Me)[2],Thr[4], Orn[8],des-Gly-$NH_2{}^9$)-vasotocin, 1 μM) blocked the TGOT-induced effects, indicating that these actions were solely mediated by the oxytocin receptor[10]. The TGOT-induced increase in spontaneous IPSCs was also abolished by the potent P/Q-type calcium channel blocker ω-Agatoxin IVA, but unaffected by the N-type calcium channel antagonist ω-Conotoxin

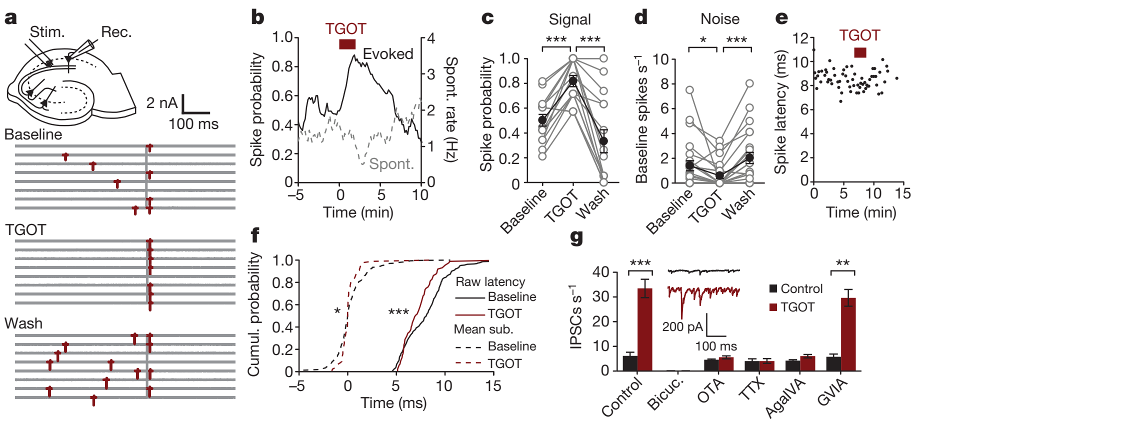


Figure 1 | Oxytocin receptor agonist (TGOT) reduces spontaneous firing but enhances EPSP-spike coupling in CA1 pyramidal neurons. **a**, Exemplar CA1 pyramidal cell-attached recording of spikes (red). Vertical bar indicates Schaffer Collateral stimulus. **b**, Time course and **c**, **d**, group data of evoked spike probability ($n = 15$ cells) and spontaneous activity ($n = 23$ cells) in pyramidal cell-attached recordings as influenced by 200 nM TGOT. **e**, Exemplar recording demonstrating TGOT reduction in evoked spike latency and jitter. **f**, Cumulative distribution of raw (solid) and mean-subtracted (dashed) spike times ($n = 15$ cells). Mean latency: 5.87 ± 0.42 ms; mean variance: $2.29 \pm 0.30\ \mathrm{ms}^2$. **g**, Spontaneous IPSC frequency onto CA1 pyramidal cells ($n = 6$ cells, each condition). Paired two-tailed t-test in panels **c**, **d** and **g**. Two-sample Kolmogorov-Smirnov (K-S) test in panel **f**. $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. Error bars, s.e.m.

[1]Department of Molecular and Cellular Physiology, 279 Campus Drive, Stanford University School of Medicine, Stanford, California 94305, USA. [2]NYU Neuroscience Institute, New York University, 450 East 29th Street, New York, New York 10016, USA.

GVIA, indicating that these events probably arise primarily from fast-spiking interneurons (FSIs) with little contribution from regular-spiking (RS) interneurons (Fig. 1g, Supplementary Fig. 1)[11,12].

To test more directly whether TGOT precisely targeted FSI subtypes, we used whole cell recordings in CA1 strata oriens and pyramidale, as stratum radiatum interneurons are unresponsive to TGOT[10] and lack OXTR expression[13]. We found a clear distinction: FSIs were responsive to TGOT, whereas RS interneurons were not (Fig. 2a). FSIs displayed robust responses upon application of 20 and 200 nM TGOT (Supplementary Fig. 2a–c), the latter producing a near-saturated effect. Dividing the increase in IPSCs onto pyramidal cells (27.3 Hz, Fig. 1g) by the increase in FSI firing rate (8.8 Hz per FSI, Fig. 2a), we calculate that, on average, each pyramidal cell receives input from at least ~3.1 TGOT-responsive FSIs in our slices.

To clarify mechanisms by which TGOT depolarizes FSIs, we voltage clamped FS perisomatic-targeting (basket and axo-axonic) and RS basket cells at −65 mV. TGOT induced a large inward current in FSIs (Fig. 2b, Supplementary Fig. 2g), but as expected had no effect on the RS cells (data not shown). TGOT also increased the rate of spontaneous inhibitory postsynaptic currents (IPSCs) onto FSIs (Supplementary Fig. 2d–f), as predicted from the FSI–FSI connectivity that may serve to regulate the distribution and extent of inhibition.

To test whether the TGOT-induced inward current arises from G protein signalling within the FSI itself, we replaced the GTP in the intracellular recording solution with 1 mM GTPγS, a non-hydrolysable GTP analogue that renders G proteins constitutively active. Action of GTPγS in inducing inward current largely occluded the effect of TGOT (Fig. 2b, Supplementary Fig. 2g), verifying that the TGOT effects involve G protein signalling within the recorded neuron. The amplitude and kinetics of the TGOT-induced current were unaffected by intracellular BAPTA, indicating that the intracellular signalling mechanism is probably not Ca^{2+}-dependent[14]. In voltage ramp recordings from FSIs, the TGOT-induced current reversed at −3.1 ± 3.4 mV (Fig. 2c, Supplementary Fig 2h), suggesting that the currents are generated by a non-selective cation channel. Partial replacement of external sodium by NMDG (50 mM) shifted the reversal potential to more negative values (−13.8 ± 3.7 mV, $P < 0.05$ unpaired two-tailed t-test, data not shown), pointing to Na^+ as the predominant charge carrier of the TGOT-induced inward current.

To investigate the mechanisms of the enhanced fidelity of spike transmission, we obtained whole cell current clamp recordings from CA1 pyramidal cells and elicited spikes synaptically or by current injection on interleaved trials (Fig. 2d, e, Supplementary Figures 3 and 4a). TGOT increased the fidelity of synaptically evoked spikes in whole cell mode, paralleling its effect in cell-attached recordings (Fig. 1), but reduced the probability of evoking spikes by whole cell current injection. This apparent reduction in pyramidal cell excitability was coupled to a hyperpolarization of the cell membrane (Supplementary Fig. 4b). As TGOT had no effect on the holding current or membrane resistance in voltage clamp recordings of pyramidal neurons in the presence of bicuculline (Supplementary Fig. 4c), we concluded that the reduction in spontaneous activity and excitability was wholly attributable to enhanced inhibitory tone. This increase in inhibitory tone, however, made the enhanced EPSP-spike coupling all the more surprising.

We speculated that the enhanced EPSP-spike coupling might arise from a shift in the synaptic excitatory–inhibitory balance. Indeed, the disynaptic IPSP was reduced by TGOT (Supplementary Fig. 4g, h), and bicuculline abolished the TGOT-induced increase in evoked spike probability in cell-attached recordings (Fig. 2f). To most rigorously isolate inhibitory inputs, we stimulated the Schaffer Collateral pathway while holding the cell at 0 mV under voltage clamp, and found that the evoked disynaptic IPSC was reduced by TGOT (Fig. 2g, h). In contrast, the evoked excitatory postsynaptic current (EPSC), isolated at −65 mV in the presence of bicuculline, was unaffected. This selective reduction of the evoked IPSC while sparing the EPSC, shifts the excitatory–inhibitory (E–I) balance and could account for the increase in evoked spike probability. This reduction in feed-forward inhibition could arise from either a reduction in excitatory to inhibitory (E→I) transmission, causing fewer interneurons to be activated, or a reduction in inhibitory

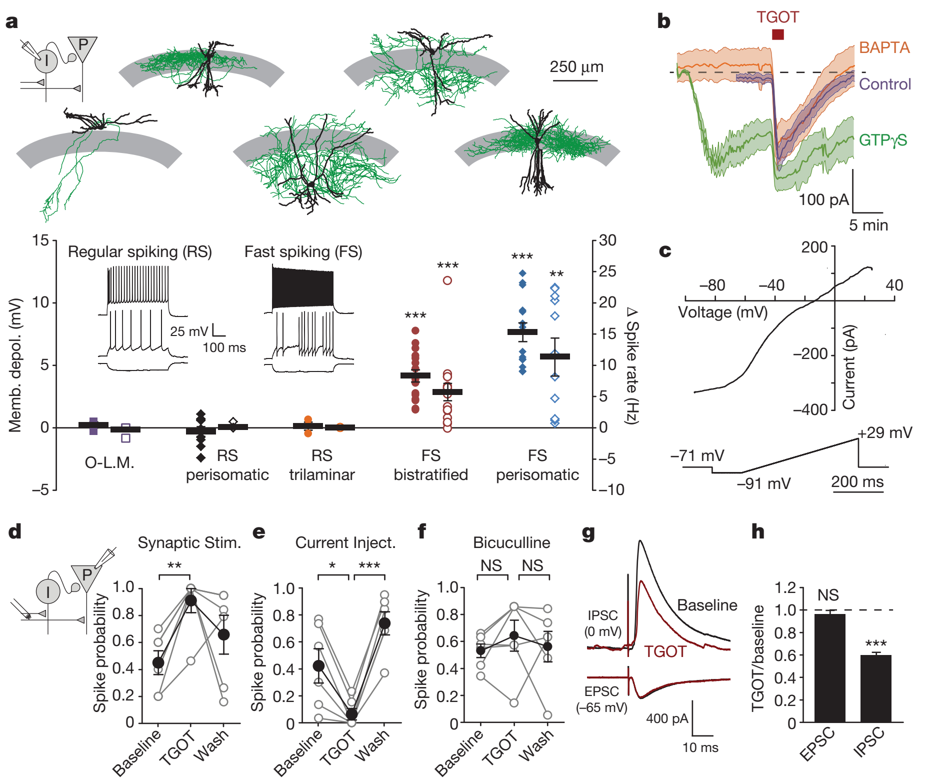


Figure 2 | TGOT activates FSIs and suppresses feed-forward inhibition. **a**, TGOT influence on membrane potential (filled symbols) and firing rate (open symbols). Exemplar electrophysiological identification of interneurons, inset. Above, exemplar biocytin-filled interneurons tracings (soma and dendrites, black; axon, green; stratum pyramidale, grey area). Oriens-lacunosum moleculare (O-L.M.; $n = 3$), RS perisomatic-targeting ($n = 9$), RS trilaminar ($n = 3$), FS bistratified ($n = 16$) and FS perisomatic-targeting ($n = 14$). **b**, Voltage clamp recordings of average TGOT currents ± s.e.m (shaded) in FSIs with control, 10 mM BAPTA and 1 mM GTPγS filled pipettes ($n = 7$, each condition). **c**, Voltage-dependence of TGOT-induced current in exemplar FSI. **d**, **e**, Whole cell spike probability from synaptically (**d**) or current injection-evoked (**e**) spikes in the same set of pyramidal neurons ($n = 6$ cells). **f**, Cell-attached, synaptically-evoked spike probability in bicuculline ($n = 6$ cells, probability: $P > 0.95$; latency: $P > 0.5$; latency variance: $P > 0.3$). Reduced stimulus strength was sufficient to reach 50% spike transmission and spikes occurred at longer latencies and with more jitter than in control ACSF[8]. **g**, TGOT influence on average evoked disynaptic IPSC from one pyramidal cell and monosynaptic EPSC from a different pyramidal cell. **h**, Normalized group data for evoked EPSC ($n = 6$ cells) and disynaptic IPSC ($n = 8$ cells). Paired two-tailed t-test, all panels except panel **d**, which uses one-tailed t-test for compatibility with cell-attached results. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Error bars, s.e.m.

to excitatory (I→E) transmission, causing each interneuron to be less effective. We recorded from FSIs while stimulating the Schaffer Collateral pathway but found no effect of TGOT on E→I transmission (Supplementary Fig. 4i, j). In contrast, stimulating the stratum pyramidale while blocking excitatory transmission with the AMPA and NMDA receptor antagonists NBQX and AP5 revealed a TGOT-induced suppression of I→E transmission onto pyramidal cells (Supplementary Fig. 4k, l).

Using paired whole cell recordings, we investigated how TGOT reduces the evoked IPSC at the I→E synapse. TGOT increased the spontaneous firing of presynaptic FSIs and also diminished the FSI–pyramidal cell unitary IPSC, without affecting RS interneurons (Fig. 3a, b). When the TGOT-induced depolarization of the presynaptic FSI was countered with a hyperpolarizing bias current, however, the spontaneous firing in the presynaptic cell ceased and the unitary evoked IPSC was maintained at its pre-TGOT amplitude. This rescue suggests that TGOT induces a use-dependent depression of the IPSC[15,16], and that the increase in spontaneous FSI firing is necessary for the reduction in the evoked feed-forward IPSC.

To test whether the TGOT-induced increase in the FSI firing rate was sufficient to account for the observed synaptic depression, we drove trains of action potentials 10 s in duration in the absence of TGOT (Fig. 3c). The frequency dependence of the residual IPSC following a 10 s train in control artificial cerebrospinal fluid (ACSF) (Fig. 3d, coloured diamonds) matched closely with that of the residual IPSC in TGOT (Fig. 3d, black symbols). Thus, the TGOT-mediated increase in FSI spontaneous firing is not only necessary (Fig. 3a, b), but also sufficient (Fig. 3c, d) to account for the observed decrease in evoked IPSC amplitude (Fig. 2g, h), and enhancement of EPSP-spike coupling (Fig. 1). Recovery of the IPSC was nearly complete by 4.5 s following the 50 Hz train, consistent with a rapid switching of the FS synapses between baseline and depressed states[16] (Fig. 3e). We also observed a modest, frequency-dependent increase in the spike width over the 10 s trains (Supplementary Fig. 5) that would be expected, if anything, to increase presynaptic release, contrary to the depression that was observed.

The specificity of TGOT for FSIs suggested that this mechanism may be a general property of this network (Supplementary Fig. 6a), and that any peptide, network state, or signal that increases the spontaneous activity of FSIs will also increase the fidelity of spike transmission. We tested this hypothesis using two independent approaches, first stimulating FSIs with the peptide cholecystokinin (CCK), and second, targeting this population with the light-activated ion channel channelrhodopsin-2 (ChR2).

CCK activates FS basket cells[11], transiently increasing their firing rate in a manner reminiscent of TGOT. In close agreement with our TGOT results, CCK enhanced inhibitory tone and suppressed the evoked feed-forward IPSC without affecting the evoked EPSC (Fig. 4a, b, Supplementary Fig. 6b–e). In cell-attached recordings, CCK increased the probability of evoking spikes in CA1 pyramidal cells by Schaffer Collateral stimulation, while simultaneously suppressing the spontaneous firing of these cells (Fig. 4c, d). Furthermore, both the latency and the jitter of the evoked spikes were reduced by CCK (Supplementary Fig. 6f), just as they were with TGOT (Fig. 1f).

We then used ChR2 to selectively activate FSIs in acute hippocampal slices from PV-Cre BAC transgenic mice. Immunostaining confirmed that the ChR2 was efficiently targeted to the parvalbumin-expressing (PV^+) FSIs (Supplementary Fig. 7). Optogenetic activation of FSIs induced IPSCs in CA1 pyramidal cells that showed a strong synaptic depression (Fig. 4e), consistent with our paired recording data (Fig. 3), and with previous reports[15,16]. In agreement with our TGOT and CCK results, driving FSIs with a brief train of blue light pulses preceding an electrical stimulus to the Schaffer Collateral pathway modestly increased the probability of eliciting a spike in a postsynaptic pyramidal cell, relative to interleaved control trials in which the blue light was omitted (Fig. 4f, Supplementary Fig. 8a). Examination of the spike latency and jitter, however, revealed a high probability that monosynaptic inhibition contaminated a subset of these recordings. When the recordings with the shortest latency and lowest jitter were excluded (Supplementary Fig. 8d–f, see Methods), the remaining neurons all exhibited a pronounced increase in evoked spike probability following ChR2 stimulation (Fig. 4f, g). Taken together, three interventions, TGOT, CCK and ChR2, therefore all converge on a single surprising conclusion: that activation of FSIs enhances the fidelity of spike transmission in the hippocampus. TGOT also increased the evoked population spike amplitude in the presence of kainate-induced gamma rhythms (Supplementary Fig. 9), thus confirming that the TGOT-induced enhancement in EPSP-spike coupling is robust under more *in vivo*-like conditions.

We constructed a minimal computational model to investigate the mechanisms linking FSI activation to the evoked spike probability,

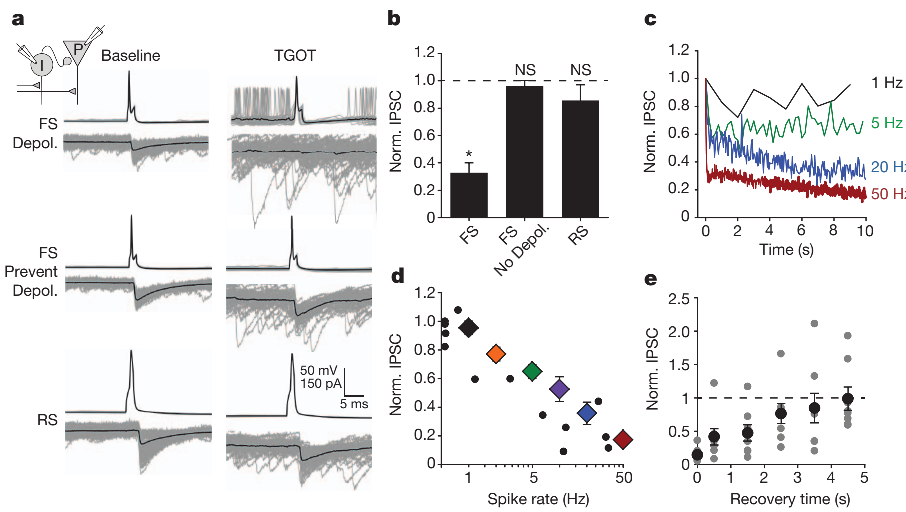


Figure 3 | Paired recordings reveal synaptic locus of TGOT-induced decrease in evoked inhibition. **a**, Presynaptic interneurons (upper) and postsynaptic pyramidal cells (lower). Individual sweeps in grey, average in black. Presynaptic FSI permitted to depolarize in TGOT (top). TGOT depolarization of presynaptic FSI countered by current injection (middle). RS interneuron transmission unaffected by TGOT (bottom). **b**, FSI–pyramidal synapses depress only when FSI is depolarized by TGOT. FSI, $n = 8$; no depolarization, $n = 5$; RS, $n = 4$. **c**, Frequency-dependent depression of FSI–pyramidal cell synapses in control ACSF. Averages normalized to first IPSC (2 Hz, 10 Hz, $n = 7$ cell pairs; 1 Hz, 5 Hz, 20 Hz, 50 Hz, $n = 8$ cell pairs). **d**, FSI–pyramidal synaptic depression from TGOT-induced firing (black circles, one point per cell pair) matches depression by 10 s spike trains in control ACSF (coloured diamonds, average from multiple cell pairs). No black circles obscured by coloured diamonds. **e**, Synaptic recovery following 50 Hz, 10 s train. ($n = 8$ cells). Paired two-tailed t-test. $*P < 0.05$. NS, not significant $P > 0.15$. Error bars, s.e.m.

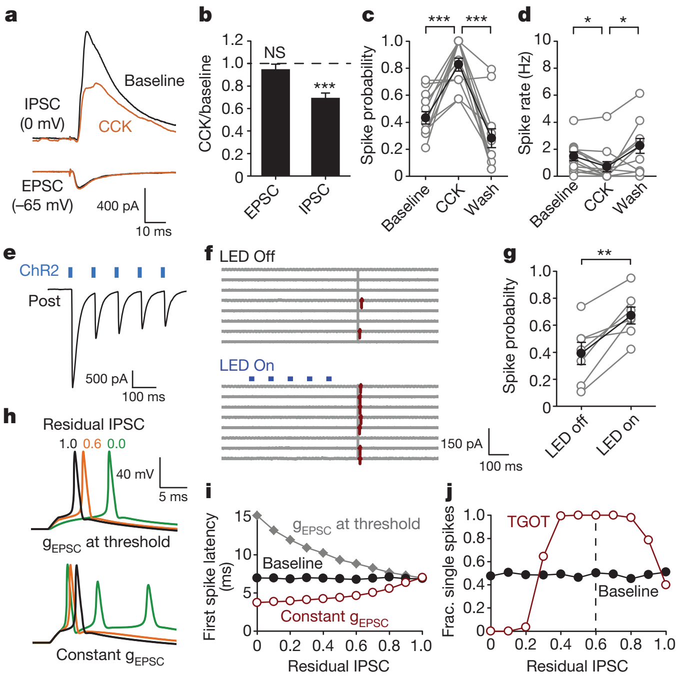


Figure 4 | Generalization to other brain states and modulators. **a**, CCK (200 nM) influence on average evoked disynaptic IPSC from one pyramidal cell and monosynaptic EPSC from a different pyramidal cell. **b**, Normalized group data for evoked EPSC ($n = 6$ cells) and disynaptic IPSC ($n = 6$ cells). **c**, Evoked spike probability ($n = 14$ cells) and **d**, spontaneous firing rate ($n = 14$ cells) in cell-attached recordings as influenced by CCK. **e**, Exemplar ChR2-evoked IPSCs recorded in a CA1 pyramidal neuron from a PV-Cre mouse injected with double-floxed AAV-ChR2. **f**, Cell-attached recording from mouse CA1 pyramidal neuron. Control (LED off) and ChR2 stimulation (LED on) sweeps interleaved during recording, but grouped for presentation. **g**, ChR2 influence on cell-attached spike probability in the subset of mouse pyramidal neurons in which latency and jitter indicated a minimal monosynaptic inhibition (see Methods). **h**, Computer simulated exemplar traces in which IPSC conductance (g_{IPSC}) is reduced and EPSC conductance (g_{EPSC}) is either lowered to maintain ~50% chance of spiking (top) or g_{EPSC} is held constant (bottom). **i**, Residual IPSC influence on simulated spike latency with g_{EPSC} held constant (red) or reduced to maintain ~50% spike probability (grey). Panels **h** and **i** generated in absence of spontaneous IPSCs to isolate feed-forward IPSC contribution to evoked spike timing. **j**, Residual IPSC influence on probability of eliciting exactly one spike (g_{EPSC} held constant). Paired two-tailed *t*-test. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$. Error bars, s.e.m.

latency and jitter. We mimicked the enhanced FSI activity by increasing the rate and amplitude of spontaneous IPSCs. The synaptic depression at the FSI–pyramidal synapse was simulated by reducing the evoked IPSC to 60% of its basal value. In agreement with our experimental results, these changes reduced the simulated evoked IPSP, increased the simulated evoked spike probability, and sharpened the evoked spike timing (Supplementary Fig. 10).

We then asked why a decrease in feed-forward inhibition shrinks evoked spike latency and jitter, in apparent conflict with the idea that feed-forward inhibition enforces sharp spike timing[8,9]. Resolution is achieved by considering how the EPSC and IPSC conductances (g_{EPSC} and g_{IPSC}) regulate membrane voltage near the spike-firing threshold. A reduction in g_{IPSC} allows an unaltered g_{EPSC} to push the membrane potential up to the spike firing threshold more reliably and more quickly and precisely (Fig. 4h, i). In contrast, if g_{EPSC} is reduced to nearly the same degree as g_{IPSC} in order to clamp the likelihood of spike firing[8], the latency and jitter are increased.

Finally, we probed the functional consequences of the strikingly incomplete depression of the FS synapses (Fig. 3c) and the effects of varying the latency between the onset of g_{EPSC} and g_{IPSC}. Fidelity of spike transmission (defined as the fraction of sweeps containing precisely one postsynaptic spike) is maximal when IPSCs are depressed by approximately 40% (Fig. 4j), the value we observed experimentally in response to either TGOT or CCK application (Figs 2g, h and 4a, b). Likewise, a residual IPSC of 50–60% was optimal in considerations of global spike jitter (Supplementary Fig. 10h). Thus, the empirically observed TGOT response in FSIs seems well suited in multiple respects to the efficient retuning of overall circuit performance.

Our experiments reveal a generalized mechanism through which oxytocin improves the fidelity and temporal precision of information transfer through brain networks. Oxytocin enhanced circuit performance in three ways: increasing throughput of output spikes, sharpening submillisecond spike timing, and suppressing background firing. Each of these improvements in circuit signal-to-noise ratio could be traced to the action of oxytocin on FSIs and reproduced through quantitative simulations, as well as through other interventions that specifically activate FSIs.

The rapid onset and recovery of FSI use-dependent synaptic depression is well suited to shift circuit dynamics rapidly yet stably in response to oxytocin, whether delivered quickly and focally, as in synaptic release[4], or presented diffusely at low doses, as in volume transmission. The partial depression of FSI synapses (residual, ~35%) and the sparing of RS interneurons ensures that modulation by oxytocin avoids the dangers associated with a complete loss of inhibition such as dramatically impaired spike timing precision[8] (Supplementary Figs 10 and 11) and epileptogenesis.

Our experiments provide a circuit mechanism linking three disparate aspects of ASD[17]. Oxytocin signalling has been implicated in ASD by genetic analysis and pharmacological studies[5,6,18,19]. PV-positive FSIs are important in autism aetiology[20], presumably due to their role in excitation–inhibition balance and neuronal oscillations, both of which are likely impaired in ASD. Deficiencies in signal-to-noise ratio, observed as unreliable cortical evoked potentials in ASD[21], offer a valuable endophenotype, but have not yet been linked to a circuit defect or a therapeutic strategy. Tying these aspects together, our finding that FSIs are direct targets of oxytocin and can potently modulate circuit signal-to-noise ratio, shows these cells may be uniquely poised to counteract deficits in rapid information processing in psychiatric disorders[18,19,21]. In healthy individuals, oxytocin signalling through FSIs may provide a salience cue, capable of transiently enhancing cognitive performance[1,3,22]. Indeed, increasing PV^+ interneuron activity was sufficient to recover hippocampal-dependent behavioural deficits in a mouse model of Alzheimer's disease[23]. There may be parallels in the visual cortex as well, where optogenetic activation of PV^+ interneurons operates like a salience cue and sharpens orientation tuning[24].

The selective action of oxytocin on FSIs, amidst the wide variety of interneuron types, raises questions about functional logic. Specific targeting of FSIs may be geared toward altering network function through fine-tuning of feed-forward inhibition. Importantly, the FSIs engaged by oxytocin are physiologically and functionally distinct from RS interneurons, which play a major role in feed-back inhibition and whose output is regulated by endocannabinoids[25]. By selectively targeting distinct interneuron populations, neuromodulators like oxytocin and endocannabinoids could be specialized for sculpting different forms of inhibition.

Another modulator, noradrenaline, enhances circuit signal-to-noise ratio in slice and *in vivo* through a variety of mechanisms across multiple brain regions including the hippocampus[26,27] and auditory system[28]. In auditory brainstem, Kuo and Trussell described how noradrenaline suppresses cartwheel inhibitory neuron spiking, relieving their output synapses from tonic depression[28]. Although this mechanism differs from ours in direction of change and functional outcome, an emergent general principle is that modulation of inhibitory neuron tonic firing and variation in use-dependent synaptic depression can regulate signal-to-noise. In the hippocampus, several monoamine responses have been delineated across excitatory and inhibitory neurons that enhance circuit signal-to-noise ratio[26,29,30]. Although oxytocin and

noradrenaline both enhance the signal-to-noise ratio of CA1 pyramidal neurons, the widespread effects of noradrenaline contrast sharply with the exquisitely focused mechanism we uncovered. Oxytocin accomplishes both the enhanced fidelity of spike transmission and the suppression of background activity by selectively targeting a single locus: FSI activity. Furthermore, FSI synaptic depression in hippocampal CA1 (Fig. 3) is representative of that in dentate gyrus[16], cortex[15], and elsewhere, indicating that similar modulation of signal-to-noise ratio by FSI activity may be essential in many brain regions.

METHODS SUMMARY

A full description of materials and Methods including slicing procedure, recording methodology, drugs, reagents, mouse lines, viruses, interneuron labelling and classification and computer modelling is available in the Supplementary Information. Briefly, acute hippocampal slices (350-μm thick) were prepared from Sprague-Dawley rats aged p21–p28 of either gender. Gender of animals did not significantly influence the effect of TGOT on EPSP-spike coupling or on interneuron depolarization (Supplementary Fig. 13). For optogenetics experiments (Fig. 4), acute hippocampal slices (300-μm thick) were prepared for recording from PV-Cre mice 3–5 weeks following virus injection. Modelling was performed using NEURON (http://www.neuron.yale.edu/neuron/) and MATLAB. All protocols were approved by the Institutional Animal Care and Use Committee of Stanford University.

Full Methods and any associated references are available in the online version of the paper.

Received 22 January; accepted 31 May 2013.
Published online 4 August 2013.

1. Domes, G., Heinrichs, M., Michel, A., Berger, C. & Herpertz, S. C. Oxytocin improves "mind-reading" in humans. *Biol. Psychiatry* **61,** 731–733 (2007).
2. Ferguson, J. N., Aldag, J. M., Insel, T. R. & Young, L. J. Oxytocin in the medial amygdala is essential for social recognition in the mouse. *J. Neurosci.* **21,** 8278–8285 (2001).
3. Tomizawa, K. *et al.* Oxytocin improves long-lasting spatial memory during motherhood through MAP kinase cascade. *Nature Neurosci.* **6,** 384–390 (2003).
4. Knobloch, H. S. *et al.* Evoked axonal oxytocin release in the central amygdala attenuates fear response. *Neuron* **73,** 553–566 (2012).
5. Modahl, C. *et al.* Plasma oxytocin levels in autistic children. *Biol. Psychiatry* **43,** 270–277 (1998).
6. Wu, S. *et al.* Positive association of the oxytocin receptor gene (*OXTR*) with autism in the Chinese Han population. *Biol. Psychiatry* **58,** 74–77 (2005).
7. Mühlethaler, M., Charpak, S. & Dreifuss, J. J. Contrasting effects of neurohypophysial peptides on pyramidal and non-pyramidal neurones in the rat hippocampus. *Brain Res.* **308,** 97–107 (1984).
8. Pouille, F. & Scanziani, M. Enforcement of temporal fidelity in pyramidal cells by somatic feed-forward inhibition. *Science* **293,** 1159–1163 (2001).
9. Buzsáki, G. & Eidelberg, E. Commissural projection to the dentate gyrus of the rat: evidence for feed-forward inhibition. *Brain Res.* **230,** 346–350 (1981).
10. Zaninetti, M. & Raggenbass, M. Oxytocin receptor agonists enhance inhibitory synaptic transmission in the rat hippocampus by activating interneurons in stratum pyramidale. *Eur. J. Neurosci.* **12,** 3975–3984 (2000).
11. Földy, C., Lee, S. Y., Szabadics, J., Neu, A. & Soltesz, I. Cell type-specific gating of perisomatic inhibition by cholecystokinin. *Nature Neurosci.* **10,** 1128–1130 (2007).
12. Wilson, R. I., Kunos, G. & Nicoll, R. A. Presynaptic specificity of endocannabinoid signaling in the hippocampus. *Neuron* **31,** 453–462 (2001).
13. Yoshida, M. *et al.* Evidence that oxytocin exerts anxiolytic effects via oxytocin receptor expressed in serotonergic neurons in mice. *J. Neurosci.* **29,** 2259–2271 (2009).
14. Alberi, S., Dreifuss, J. J. & Raggenbass, M. The oxytocin-induced inward current in vagal neurons of the rat is mediated by G protein activation but not by an increase in the intracellular calcium concentration. *Eur. J. Neurosci.* **9,** 2605–2612 (1997).
15. Galarreta, M. & Hestrin, S. Frequency-dependent synaptic depression and the balance of excitation and inhibition in the neocortex. *Nature Neurosci.* **1,** 587–594 (1998).
16. Kraushaar, U. & Jonas, P. Efficacy and stability of quantal GABA release at a hippocampal interneuron-principal neuron synapse. *J. Neurosci.* **20,** 5594–5607 (2000).
17. Persico, A. M. & Bourgeron, T. Searching for ways out of the autism maze: genetic, epigenetic and environmental clues. *Trends Neurosci.* **29,** 349–358 (2006).
18. Guastella, A. J. *et al.* Intranasal oxytocin improves emotion recognition for youth with autism spectrum disorders. *Biol. Psychiatry* **67,** 692–694 (2010).
19. Anagnostou, E. *et al.* Intranasal oxytocin versus placebo in the treatment of adults with autism spectrum disorders: a randomized controlled trial. *Mol. Autism* **3,** 16 (2012).
20. Gogolla, N. *et al.* Common circuit defect of excitatory-inhibitory balance in mouse models of autism. *J. Neurodevelop. Disord.* **1,** 172–181 (2009).
21. Dinstein, I. *et al.* Unreliable evoked responses in autism. *Neuron* **75,** 981–991 (2012).
22. Kinsley, C. H. *et al.* Motherhood improves learning and memory. *Nature* **401,** 137–138 (1999).
23. Verret, L. *et al.* Inhibitory interneuron deficit links altered network activity and cognitive dysfunction in Alzheimer model. *Cell* **149,** 708–721 (2012).
24. Lee, S. H. *et al.* Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* **488,** 379–383 (2012).
25. Glickfeld, L. L. & Scanziani, M. Distinct timing in the activity of cannabinoid-sensitive and cannabinoid-insensitive basket cells. *Nature Neurosci.* **9,** 807–815 (2006).
26. Madison, D. V. & Nicoll, R. A. Actions of noradrenaline recorded intracellularly in rat hippocampal CA1 pyramidal neurones, *in vitro*. *J. Physiol. (Lond.)* **372,** 221–244 (1986).
27. Segal, M. & Bloom, F. E. The action of norepinephrine in the rat hippocampus. IV. The effects of locus coeruleus stimulation on evoked hippocampal unit activity. *Brain Res.* **107,** 513–525 (1976).
28. Kuo, S. P. & Trussell, L. O. Spontaneous spiking and synaptic depression underlie noradrenergic control of feed-forward inhibition. *Neuron* **71,** 306–318 (2011).
29. Bergles, D. E., Doze, V. A., Madison, D. V. & Smith, S. J. Excitatory actions of norepinephrine on multiple classes of hippocampal CA1 interneurons. *J. Neurosci.* **16,** 572–585 (1996).
30. Doze, V. A., Cohen, G. A. & Madison, D. V. Synaptic localization of adrenergic disinhibition in the rat hippocampus. *Neuron* **6,** 889–900 (1991).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Földy for generous assistance with techniques for interneuron recording and classification, and thank Y. Li, M. Tadross, A. Mitra, D. Poburko, L. Prolo, J. Huguenard, S. Hestrin, D. Madison, L. Luo, K. Pelkey and K. Deisseroth for discussion and comments. This work was supported by the Burnett Family Fund, the Mosbacher Fund, and the Mathers Foundation, as well as grants from the National Institute of Mental Health (MH064070, MH071739), the National Institute of Neurological Disorders and Stroke (NS024067) and the Simons Foundation (95395). S.F.O. is supported by an NDSEG fellowship and by an NIMH-NRSA fellowship (F31MH084430). P.L.B. is supported by an SNSF and SFGBM fellowship (PASMP3_140063/1).

Author Contributions S.F.O. performed all experiments, analysis and modelling. S.N.T. and S.F.O. performed ChR2 experiments. P.L.B. and N.N.T. assisted with interpretation of results and long-term project direction. R.W.T. and G.F. oversaw experiments, analysis, modelling and project direction. S.F.O. and R.W.T. planned and designed experiments and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.W.T. (richard.tsien@nyumc.org).

METHODS

Slice preparation. Rat hippocampal slices (350-μm thick) were prepared using a Leica VT 1000S vibratome from p21–p28 Sprague-Dawley rats of either sex in ice-cold sucrose slicing solution containing (in mM) 206 Sucrose, 11 D-Glucose, 2.5 KCl, 1 NaH_2PO_4, 10 $MgCl_2$, 2 $CaCl_2$ and 26 $NaHCO_3$. Rats were anaesthetized with isofluorane inhalation before decapitation and dissecting out of the hippocampus. Mouse transverse hippocampal slices (300-μm) were prepared using a Vibratome 1000 plus (Vibratome). Mice were deeply anaesthetized with intraperitoneal injection of pentobarbital (100 mg per kg body weight) and then transcardially perfused with ~30 ml ice-cold sucrose-ACSF solution containing (in mM) 252 sucrose, 24 $NaHCO_3$, 1.25 NaH_2PO_4, 3 KCl, 2 $MgSO_4$, 10 D-Glucose and 0.5 $CaCl_2$. All slices from rats and mice were allowed to recover submerged in artificial cerebrospinal fluid (ACSF) for 1 h at 34 °C, and then maintained at room temperature until recording. For recordings from rat tissue, ACSF contained (in mM) 122 NaCl, 3 KCl, 10 D-glucose, 1.25 NaH_2PO_4, 2 $CaCl_2$, 1.3 $MgCl_2$, 26 $NaHCO_3$, 3 sodium pyruvate, 2 sodium ascorbate and 5 L-glutamine. For mouse recordings, ACSF contained (in mM) 124 NaCl, 26 $NaHCO_3$, 2.5 KCl, 1.25 NaH_2PO_4, 2 $CaCl_2$, 2 $MgSO_4$, 5 L-Glutamine, and 10 D-Glucose. All slice preparation and recording solutions were oxygenated with carbogen gas (95% O_2, 5% CO_2, pH 7.4).

Electrophysiological recordings. Recordings were performed in a submerged chamber at 32–34 °C with constant bath perfusion of ACSF at ~5 ml minute^{-1} for rats, ~2 ml minute^{-1} for mice. Slices were allowed 15–45 min to equilibrate before recording. Because the $GABA_B$ blocker CGP52432 (2 μM) did not affect the TGOT enhancement of evoked spike probability, recordings were pooled from control ACSF ($n = 7$ cells) and CGP52432 ($n = 8$ cells) conditions to measure spike probability, suppression of spontaneous firing, and evoked spike timing (Fig. 1). For cell-attached measurement of TGOT influence on spontaneous activity, results were pooled from recordings in control ACSF ($n = 15$ cells) and in the presence of CGP52432 at 2 μM ($n = 8$ cells). To prevent ictal activity, the CA3 region of each slice was removed before recordings in bicuculline. Recordings were made using glass pipettes with a tip resistance of 2–4 MΩ. For cell-attached recordings, pipettes were filled with ACSF and the amplifier was set in voltage clamp mode. Slices were visualized with an upright microscope (Zeiss Axioskop 2 FS plus) using infrared differential interference contrast (IR-DIC) optics. Data were recorded with a MultiClamp 700B amplifier (Axon Instruments), filtered at 10 kHz using a Bessel filter and digitized at 20 kHz with a Digidata 1322A analogue-digital interface (Axon Instruments). For whole cell recordings, experiments were discarded if the series resistance changed significantly or reached 20 MΩ. Spontaneous IPSCs onto pyramidal cells and unitary IPSCs in paired recordings were detected in voltage clamp using a high Cl^- internal solution containing (in mM) 70 $CsMeSO_3$, 35 CsCl, 15 TEA-Cl, 1 $MgCl_2$, 0.2 $CaCl_2$, 10 HEPES, 0.3 EGTA, 10 Tris-phosphocreatine, 4 Mg-ATP, and 0.3 Na-GTP. For evoked IPSC and EPSC recordings, the internal solution contained (in mM) 130 $CsMeSO_3$, 8 CsCl, 1 $MgCl_2$, 10 HEPES, 0.3 EGTA, 10 Tris-phosphocreatine, 4 Mg-ATP, and 0.3 Na-GTP. Bicuculline (10 μM), TTX (100 nM) and OTA (1 μM) were delivered as indicated in the bathing solution throughout the recording (Fig. 1, Supplementary Figure 1). Calcium channel blockers ω-agatoxin IVA at 0.5 μM (AgaIVA) or ω-conotoxin GVIA at 1 μM (GVIA) were delivered by pretreating the slice for 30 min in an interface chamber before recording in control ACSF. AgaIVA and GVIA recordings were performed in separate slices from the same experimental animal.

Synaptic events were evoked using a tungsten bipolar stimulating electrode placed in the Schaffer Collateral excitatory afferents from area CA3 to deliver stimuli 100 μs in duration. With the exception of Supplementary Fig. 4k, l, the stimulating electrode was placed far from the recorded cell (~400 μm to ~800 μm) to minimize monosynaptically evoked IPSCs. In Supplementary Fig. 4k, l, monosynaptic IPSCs were evoked using submaximal stimulation by placing the stimulating electrode in the pyramidal cell layer close to the recorded cell (~100 μM), and including 10 μM NBQX and 50 μM AP5 in the bath to block excitatory transmission. For evoked IPSP measurement, data were pooled from evoked spike successes and failures and from recordings in the presence ($n = 5$ cells) or the absence ($n = 1$ cell) of the $GABA_B$ antagonist CGP52432 (2 μM). Evoked disynaptic feed-forward IPSCs (Figs 2g, h and 4a, b) were recorded as outward currents at a holding potential of 0 mV in control ACSF. Evoked EPSCs were isolated by including 10 μM bicuculline in the bath and holding the cell at −65 mV. Two out of 14 recordings in Figs 4c, d and Supplementary Figures 6e, f were performed in the continuous presence of AM-251 (2 μM) to confirm the persistence of the CCK-induced enhancement of EPSP–spike coupling even when endocannabinoid signalling was blocked.

For current clamp recordings, and all interneuron recordings except for the voltage ramp experiments, the intracellular solution contained (in mM) 130 K-Gluconate, 1 $MgCl_2$, 10 HEPES, 0.3 EGTA, 10 Tris-phosphocreatine, 4 Mg-ATP, and 0.3 Na-GTP. For interneuron recordings this solution was supplemented with 0.1% biocytin. GTP was omitted in experiments featuring GTPγS. For voltage clamp recordings of TGOT-induced currents in FSIs, traces were divided into 10 s segments, with the mean value of each segment plotted as a function of time to exclude synaptic events. See Supplementary Fig. 2d for an exemplar raw trace. All recordings were baseline-subtracted to adjust for the leak current measured during the first 2 min before the onset of the GTPγS-induced current. Traces were time-aligned to the wash-in of TGOT (red bar). For one cell in the GTPγS data set in which baseline recording period was 10 min rather than 15 min, the pre-TGOT period was aligned to the start of the other recordings, and the remainder of the trace starting with TGOT wash-in was aligned to the TGOT wash-in of the other traces.

Low doses of kainate (100–500 nM) were used to establish gamma rhythms in hippocampal slices that closely resemble gamma rhythms *in vivo*[31]. Field recording electrodes were placed in the strata pyramidale and radiatum to monitor both the ongoing gamma oscillation and the EPSP-spike coupling. Field recording electrodes were similar to those used for whole cell recordings, but filled with ACSF.

FSI voltage ramp recordings. For voltage ramp recordings, the internal solution contained (in mM) 50 K-Gluconate, 70 $CsMeSO_3$, 10 TEA-Cl, 1 $MgCl_2$, 10 HEPES, 0.3 EGTA, 10 Tris-phosphocreatine, 4 Mg-ATP, and 0.3 Na-GTP. The pipette reference potential was set to zero and a junction potential of −15.1 mV (calculated using pClamp) was corrected post hoc. An additional, empirically measured correction factor of 3.3 mV was applied to correct for a change in the junction potential introduced by partial replacement of sodium with NMDG in the voltage ramp ACSF. Apart from the voltage ramp recordings, other membrane potentials reported are not corrected for liquid junction potentials. After obtaining a whole cell recording from a putative interneuron, the fast-spiking phenotype was verified as described below. The amplifier was then switched to voltage clamp mode and the bath solution was substituted for voltage ramp ACSF containing (in mM) 112 NaCl, 10 D-Glucose, 3 KCl, 1.25 NaH_2PO_4, 10 TEA-Cl, 1.3 $MgCl_2$, 2 $CaCl_2$,26 $NaHCO_3$, 5 4-Aminopyridine, 0.1 $CdCl_2$ and 0.001 TTX. Voltage ramps ~1 s in duration between −91 and +29 mV were applied once every 10 s until the current at each potential reached a steady state for > 2 min, at which point TGOT was applied. In 3 out of 13 recordings the voltage ramp-activated current (1) became more negative at all potentials shortly after TGOT application, and (2) failed to return to baseline after washout of the drug. It was assumed that this global shift was caused by a change in the space clamp or access resistance and these recordings were excluded from further analysis.

Drugs and reagents. All salts and buffers for intracellular and extracellular solutions, as well as ATP, GTP, GTPγS, phosphocreatine and biocytin were purchased from Sigma. TGOT ((Thr^4,Gly^7)-oxytocin), OTA ((d(CH_2)$_5$^{1},Tyr(Me)^2,Thr4,Orn^8, des-Gly-NH_2^{9})-vasotocin) and CCK (cholecystokinin octapeptide) peptides were purchased from Bachem, dissolved at 1 mM in ddH_2O and stored at −20 °C until use within 6 months of purchase. Bicuculline, TTX, NBQX and D-AP5 were purchased from Ascent Scientific. ω-conotoxin GVIA and ω-agatoxin IVA were purchased from Peptides International. Stock solutions were prepared and stored according to manufacturer specifications.

Interneuron labelling and classification. Physiological classification of interneuron subtypes was based on established criteria[11,25,32]. Fast-spiking cells were defined as those including (1) peak firing rates > 200 Hz with little firing rate accommodation, (2) characteristic FS action potential waveform, and (3) minimal hyperpolarization-induced sag current due to I_h . Following interneuron recordings, slices were transferred to a fixative solution containing 4% paraformaldehyde, 0.2% picric acid and 1× phosphate buffered saline for 24–72 h before being stained with 3,3′-diaminobenzidine tetrahydrochloride (0.015%) using a standard ABC kit (Vector). Neuronal cell types were identified based on morphology of axonal and dendritic arbors and electrophysiological properties of the cell. The FS perisomatic-targeting set includes both basket cells (shown), and axo-axonic cells (not shown). Because of technical challenges of discriminating FS basket and axo-axonic cells unequivocally, both cell types were pooled into a single group of FS perisomatic-targeting cells. When analysed separately, both putative types were equivalently responsive to TGOT.

Analysis of cell-attached and intracellular recording data. Analysis of spikes, evoked synaptic currents, and synaptic potentials were performed offline using custom written routines in MATLAB (MathWorks). Spontaneous IPSCs were detected using a modified version of the detectPSPs script by P. Larimer (http://www.mathworks.com/matlabcentral/fileexchange). Spike jitter histograms were calculated by subtracting the latency of each spike from the average latency of spikes evoked in that cell. The average latency and jitter were calculated separately for control and TGOT/CCK conditions in each cell. To measure the spike width[33], raw data was oversampled to 133 kHz using the MATLAB spline function. Time course of spontaneous activity in pyramidal cell attached recordings was calculated by averaging over all cells and smoothing in time with a boxcar filter (width = 7 sweeps).

Optical stimulation of channelrhodopsin-2. Photostimuli were produced by three Luxeon Rebel LEDs (470 nm, Philips Lumileds) driven by a custom-built controller. The LEDs were placed below the recording chamber for full slice

illumination once stable recording conditions were reached. Light pulses were 5 ms in duration with a power of approximately 0.5 mW per mm^2. ChR2-evoked IPSCs were recorded from CA1 pyramidal neurons ($n = 10$ cells, $n = 4$ animals). Six of these neurons were recorded in the same region of the same slice as neurons recorded in the cell-attached data set.

Data analysis for cell-attached recordings involving blue light stimulation. In the full data set, blue light stimulation increased the spike probability in 13 of 16 neurons (Supplementary Fig. 8a, 12% increase in spike probability including all neurons, $P < 0.05$, paired two-tailed *t*-test). In recordings from rat neurons the average increase in spike firing probability with TGOT or CCK was not correlated with spike latency, whereas in the mouse data we found a strong correlation between control spike latency and the ChR2-induced increase in spike probability (Supplementary Fig. 8d, e). In the mouse data set, the shortest latency spikes showed the weakest increase in spike firing probability. Plotting the latency against the jitter of spikes elicited under control conditions, we found a clear separation between two groups of cells, in which evoked spikes from one set of cells occurred with very short latency and little jitter and spikes from another set of cells occurred at longer latency and with more jitter (Supplementary Fig. 8f).

Because of the smaller size of the mouse brain, we found our slice angle to be less reliably transverse than in the rat preparation. As a result, the stimulating and recording electrodes were placed closer to one another in the mouse slice in order for the stimulating electrode to recruit a sufficient number of excitatory Schaffer Collateral fibres to drive an action potential in the postsynaptic CA1 pyramidal cell. This change in recording configuration unfortunately increases the probability of directly activating inhibitory fibres with the stimulating electrode and generating a monosynaptic IPSC. A well-documented set of physiological parameters, including synaptic kinetics and cell excitability[25] ensure that the physiologically relevant disynaptic IPSC arises mostly from FSIs. The monosynaptically activated IPSC, however, will arise from a less targeted subset of neurons, and therefore be less susceptible to modulation by interventions that selectively target FSIs.

In the cell-attached recording configuration, it was impossible to determine directly the relative monosynaptic and disynaptic contributions to the feed-forward IPSC. The monosynaptic IPSC relies only on a single GABAergic synapse, however, whereas the disynaptic IPSC relies on three sequential steps: (1) a glutamatergic synapse onto the interneuron, (2) the subsequent action potential in the interneuron, and finally (3) the GABAergic transmission onto the postsynaptic pyramidal cell. The monosynaptic IPSC will therefore occur with a shorter latency and less jitter than the disynaptically evoked IPSC. As a result, spikes in pyramidal cells in which the feed-forward IPSC is dominated by a monosynaptic component will be expected to occur with a shorter latency and less jitter than spikes in cells experiencing a more physiological disynaptic feed-forward IPSC. We therefore excluded the tightly clustered group of neurons with very short latency and low jitter spikes from the mouse data set ($n = 9$ cells) and analysed only the neurons in which spikes occurred with a longer latency and more jitter, consistent with disynaptic feed-forward inhibition ($n = 7$ cells). All of these remaining cells demonstrated an increase in spike firing probability following blue light stimulation (7 out of 7 cells, 28% increase in spike probability; $P < 0.01$ paired two-tailed *t*-test). In the complete data set ($n = 16$ cells) we observed a modest increase in spike latency following blue light stimulation of PV interneurons across all 16 neurons (Supplementary Fig. 8b, c). However, in the 5 out of 7 cells from the restricted data set that fired at least 5 spikes in both the control and blue light stimulation conditions, light activation of PV interneurons reduced the latency (Supplementary Fig. 8g, 10.35 ms in control, 10.07 ms following light stimulation; $P = 0.73$ paired two-tailed *t*-test) and jitter (Supplementary Fig. 8h, 16.58 ms^2 control; 11.78 ms^2 light stimulation; $P = 0.23$ paired two-tailed *t*-test) of spikes. Although this reduction in latency and jitter did not reach statistical significance, the trend is consistent with our TGOT and CCK results.

Immunohistochemistry. At the end of each ChR2 recording session, slices were fixed overnight with 4% paraformaldehyde (PFA) in a phosphate buffered saline (PBS) solution and cryoprotected by immersion in 30% sucrose PBS solution overnight at 4 °C. Tissues were embedded in Tissue Tek, frozen on dry ice, and cryosectioned at 20-μm thickness. Sections for were processed using 1.5% normal goat serum (NGS) and 0.1% Triton X-100 in all procedures except washing steps, where only PBS was used. Sections were incubated in blocking solution for 1 h, followed by incubation with the primary antibodies overnight at 4 °C. Cryostat tissue sections were stained with the primary antibodies: mouse anti-Parvalbumin (1:1,000, Sigma) and rabbit anti-DsRed (1:500, Chemicon). Secondary antibodies conjugated with Alexa Fluor dyes 488, 594 (Molecular Probes) raised from the same host used for blocking serum were applied for 1 h at room temperature. Nuclear counterstaining was performed with 100 ng ml^{-1} 4,6-diamidino-2-phenylindole (DAPI) solution in PBS for 5 min. Fluorescent images were captured using a cooled-charge coupled device (CCD) camera (Princeton Scientific Instruments) using MetaMorph software (Molecular Devices).

Virus injection. Adeno-associated virus carrying ChR2 fused to the fluorescent marker mCherry AAV2/1.EF1.dflox.hChR2(H134R)–mCherry.WPRE.hGH, (University of Pennsylvania Gene Therapy Program Vector Core) was injected bilaterally into dorsal hippocampal CA1 region of *Pvalb-cre* (PV-Cre) transgenic mice[34] (aged between postnatal days 15–19) at three sites: 2.2, 1.8 and 1.6 mm posterior from bregma; 2.4, 2.1, 1.7 mm from midline; and 1.2, 1.1, and 1 mm below the cortical surface, respectively. Animals were anaesthetized with isoflurane, mounted in a stereotactic apparatus and kept under isoflurane anaesthesia during surgery. We injected 100 nl of virus at each location over a 2 min period using a glass micropipette (tip diameter ~20 μm) attached to a Nanolitre 2000 pressure injection apparatus (World Precision Instruments). The pipette was held in place for 3 min following each injection before being completely retracted from the brain. Mice were returned to their home cage for 2–3 weeks before acute slice preparation to allow for virus expression.

Computational model of EPSP-spike coupling. The computer modelling was performed using NEURON and automated using MATLAB. A simplified pyramidal cell, consisting of a soma, a single axon and a single dendrite was initialized to starting parameters before each stimulus. Background and voltage-gated conductances were based on reported models[35,36]. Small adjustments were made to improve agreement of parameters such as cell excitability and action potential waveform between the model and experimental observations. Each sweep consisted of (1) a 'monosynaptic' EPSC onto the dendrite, (2) a 'disynaptic' feed-forward IPSC onto the soma and dendrite 2 ms after the evoked EPSC (unless otherwise specified), and (3) multiple 'spontaneous' IPSCs onto the soma with randomly distributed amplitudes and timing. To isolate the role of the feed-forward IPSC from changes in inhibitory-tone, spontaneous IPSCs were omitted in the simulation used to generate Fig. 4h, i. At the outset of each set of sweeps, the 'evoked' EPSC–IPSC amplitudes were set empirically by increasing the EPSC and IPSC conductances together with a fixed ratio of 6:1 until ~50% chance of spike propagation was reached. Experimental measurement of IPSC/EPSC ratio ranged from 2.62 to 5.20 (mean $\pm$ s.e.m. of 3.65 $\pm$ 0.28). This experimentally measured range is presumed to be an underestimate of the true ratio due to imperfect isolation of the IPSC reversal potential, causing a presumed GABAergic contribution to the measured EPSC in some cells. In the model, IPSC/EPSC ratios from 4:1 up to 6:1 showed a pronounced TGOT-induced increase in evoked spike probability, with 6:1 supporting the strongest influence of TGOT on spike timing and jitter. Variability was introduced by using pseudo-random number generation to vary independently (1) the evoked EPSC conductance, (2) the evoked IPSC conductance and (3) the spontaneous IPSC timing and amplitudes. Evoked EPSC and IPSC conductances were varied independently on each sweep according to a normal distribution centred on the empirically determined mean value, with a standard deviation that was 5% of the mean. TGOT was simulated by (1) reducing the evoked somatic IPSC conductance to 60% of 'baseline', while sparing the evoked EPSC and the dendritic IPSC, (2) doubling the spontaneous IPSC amplitude, and (3) increasing the spontaneous IPSC rate from 5 Hz to 35 Hz. The IPSC reversal potential was set at -110 mV for Supplementary Fig. 10 b, c, consistent with the calculated $GABA_A$ reversal potential in our whole cell recording conditions. For the rest of the simulations, the IPSC reversal potential was set to -90 mV, consistent with cell-attached recording conditions. The increase in evoked spike probability was robust as the $GABA_A$ reversal potential was varied from -80 mV to -120 mV (Supplementary Fig. 12), while the reduction in latency and latency jitter were decreased in magnitude but remained statistically significant as the $GABA_A$ reversal potential approached the neuron resting membrane potential.

31. Atallah, B. V. & Scanziani, M. Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition. *Neuron* **62,** 566–577 (2009).
32. Freund, T. F. & Buzsaki, G. Interneurons of the hippocampus. *Hippocampus* **6,** 347–470 (1996).
33. Shu, Y., Hasenstaub, A., Duque, A., Yu, Y. & McCormick, D. A. Modulation of intracortical synaptic potentials by presynaptic somatic membrane potential. *Nature* **441,** 761–765 (2006).
34. Hippenmeyer, S. *et al.* A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol.* **3,** e159 (2005).
35. Ascoli, G. A., Gasparini, S., Medinilla, V. & Migliore, M. Local control of postinhibitory rebound spiking in CA1 pyramidal neuron dendrites. *J. Neurosci.* **30,** 6434–6442 (2010).
36. Hao, J., Wang, X. D., Dan, Y., Poo, M. M. & Zhang, X. H. An arithmetic rule for spatial summation of excitatory and inhibitory inputs in pyramidal neurons. *Proc. Natl Acad. Sci. USA* **106,** 21906–21911 (2009).

LETTER

doi:10.1038/nature12332

Non-vesicular trafficking by a ceramide-1-phosphate transfer protein regulates eicosanoids

Dhirendra K. Simanshu[1]*, Ravi Kanth Kamlekar[2]*, Dayanjan S. Wijesinghe[3]*, Xianqiong Zou[2], Xiuhong Zhai[2], Shrawan K. Mishra[2], Julian G. Molotkovsky[4], Lucy Malinina[5], Edward H. Hinchcliffe[2], Charles E. Chalfant[3,6], Rhoderick E. Brown[2] & Dinshaw J. Patel[1]

Phosphorylated sphingolipids ceramide-1-phosphate (C1P) and sphingosine-1-phosphate (S1P) have emerged as key regulators of cell growth, survival, migration and inflammation[1–5]. C1P produced by ceramide kinase is an activator of group IVA cytosolic phospholipase $A_2\alpha$ ($cPLA_2\alpha$), the rate-limiting releaser of arachidonic acid used for pro-inflammatory eicosanoid production[3,6–9], which contributes to disease pathogenesis in asthma or airway hyper-responsiveness, cancer, atherosclerosis and thrombosis. To modulate eicosanoid action and avoid the damaging effects of chronic inflammation, cells require efficient targeting, trafficking and presentation of C1P to specific cellular sites. Vesicular trafficking is likely[10] but non-vesicular mechanisms for C1P sensing, transfer and presentation remain unexplored[11,12]. Moreover, the molecular basis for selective recognition and binding among signalling lipids with phosphate headgroups, namely C1P, phosphatidic acid or their lyso-derivatives, remains unclear. Here, a ubiquitously expressed lipid transfer protein, human GLTPD1, named here CPTP, is shown to specifically transfer C1P between membranes. Crystal structures establish C1P binding through a novel surface-localized, phosphate headgroup recognition centre connected to an interior hydrophobic pocket that adaptively expands to ensheath differing-length lipid chains using a cleft-like gating mechanism. The two-layer, α-helically-dominated 'sandwich' topology identifies CPTP as the prototype for a new glycolipid transfer protein fold[13] subfamily. CPTP resides in the cell cytosol but associates with the trans-Golgi network, nucleus and plasma membrane. RNA interference-induced CPTP depletion elevates C1P steady-state levels and alters Golgi cisternae stack morphology. The resulting C1P decrease in plasma membranes and increase in the Golgi complex stimulates $cPLA_2\alpha$ release of arachidonic acid, triggering pro-inflammatory eicosanoid generation.

During screening of the NCBI human genome database, we noted an *in silico* predicted transcript (GenBank NP_001025056.1; glycolipid transfer protein domain-containing protein 1; *GLTPD1*) encoding a protein sharing sequence identity (17%) with glycolipid transfer protein (GLTP). Although annotation indicated glycolipid binding and transport activity, Lys and Arg substitutions occurred at key positions (N52, W96) essential for sugar headgroup recognition by GLTP (Supplementary Fig. 1a; yellow highlights)[13,14]. We validated *GLTPD1* mRNA transcript expression in human tissues, finding widespread occurrence and relatively elevated transcript levels in placenta, kidney, pancreas and testis (Fig. 1a). Cloning and heterologous expression revealed that purified GLTPD1 (GenBank JN542538) transfers anthrylvinyl-C1P (AV-C1P) between 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) bilayer vesicles in a protein concentration-dependent fashion (Fig. 1b, c and Supplementary Fig. 1b, d) requiring acceptor membranes (Supplementary Fig. 1c). Testing of lipid specificity revealed no transfer of galactosylceramide (GalCer), lactosylceramide (LacCer), sphingomyelin (SM), phosphatidylcholine (PC), phosphatidic acid (PA) or ceramide (Cer) by GLTPD1 (Fig. 1b, c). Slow-down of AV-C1P transfer by potential lipid ligands (nonfluorescent) showed no competition effect by S1P (Fig. 1d and Supplementary Fig. 1e, f) and a transfer rate of ~4 C1P molecules per min per protein molecule at 37 °C. We have designated GLTPD1 as ceramide-1-phosphate transfer protein (CPTP).

The crystal structure of human CPTP and 16:0-C1P complex (1.9 Å, Supplementary Table 1) revealed a two-layered, all α-helical topology (Fig. 1e–g) homologous with the GLTP fold[13]. αN, α1 and α2 form one layer, α4, α5 and α8 form another layer, and α3, α6 and α7 localize along the periphery of the two-layer core. A positively-charged surface cavity for anchoring the lipid headgroup (Fig. 1h) extends through a

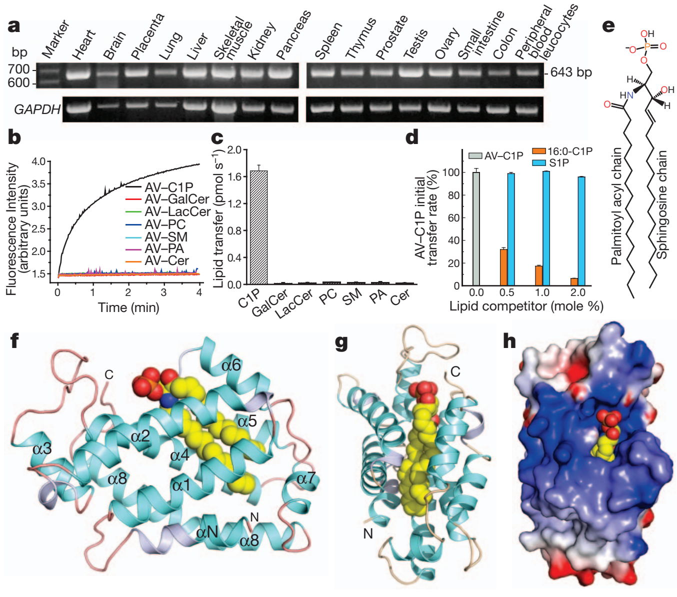


Figure 1 | CPTP lipid transfer activity and architecture. **a**, *CPTP* mRNA transcript levels in various human tissues. bp, base pairs. **b**, Lipid transfer *in vitro* by Förster resonance energy transfer. **c**, Initial lipid transfer rates for panel **b**. **d**, Competition against CPTP-mediated AV-C1P transfer by nonfluorescent lipids. Kinetic traces are shown in Supplementary Fig. 1e–g. Data in **c** and **d** represent the mean ± s.d. of three independent experiments. **e**, 16:0-C1P chemical formula. **f**, **g**, Two views of CPTP structure (ribbon) with bound 16:0-C1P (space-filling). α-helices (cyan), 3_{10}-helices (light blue), loops (orange) and bound 16:0-C1P (yellow, red, blue for carbon, oxygen, nitrogen, respectively). α-helices (αN and α1–α8) are numbered from amino (N) to carboxy (C) termini. **h**, Surface electrostatics of CPTP with bound 16:0-C1P showing positive- (blue) and negative- (red) charged residues.

[1]Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [2]Hormel Institute, University of Minnesota, Austin, Minnesota 55912, USA. [3]Department of Biochemistry and Molecular Biology, Virginia Commonwealth University, Richmond, Virginia 23298, USA. [4]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia. [5]Structural Biology Unit, CIC bioGUNE, Technology Park of Bizkaia, 48160 Derio-Bilbao, Spain. [6]Hunter Holmes McGuire Veterans Administration Medical Center, Richmond, Virginia 23249; Massey Cancer Center, Richmond, Virginia 23298, USA.
*These authors contributed equally to this work.

gateway portal, transforming into a deep interior hydrophobic cavity that accommodates the sphingosine and acyl chains. Sealing of the cavity bottom by Leu 10 and Leu 14 (αN-helix) renders it pocket-like. A triad of cationic residues (K60, R106, R110) in the surface cavity recognizes and binds the C1P phosphate headgroup (Fig. 2a). The anchoring hydrogen-bond network is complex, involving bifurcated hydrogen bonding by K60 (α2-helix) with the O1 and O2 atoms, bidentate hydrogen bonding by R106 (α4-helix) with the O2 and O3 atoms, and bidentate hydrogen bonding by R110 (α4-helix) directly and through water bridging to O3. Point mutation data supports key roles for K60 and R106 in C1P headgroup recognition with K60A and R106L showing almost no C1P transfer, whereas the R110 mutation (R110L) shows ~40% transfer (Fig. 2b). The positive-charge of this site also is enhanced by R97 (α3–4 loop) which hydrogen bonds to the O2 atom of phosphate through water-bridging. R97 is almost fully active when mutated to Leu (R97L), but mutation to acidic Glu (R97E) reduces C1P transfer to ~55%, supporting its role for attracting the lipid phosphate headgroup. Cation-pi interaction between R113 (α4-helix) and Y149 (α5–6 loop) provides stabilizing underpinning for the site (Supplementary Fig. 2c), as the R113 mutation (R113L) strongly diminishes activity. Mutants Y149A, R113E and R113L show poor C1P transfer, as expected by conformational destabilization (Fig. 2b). All

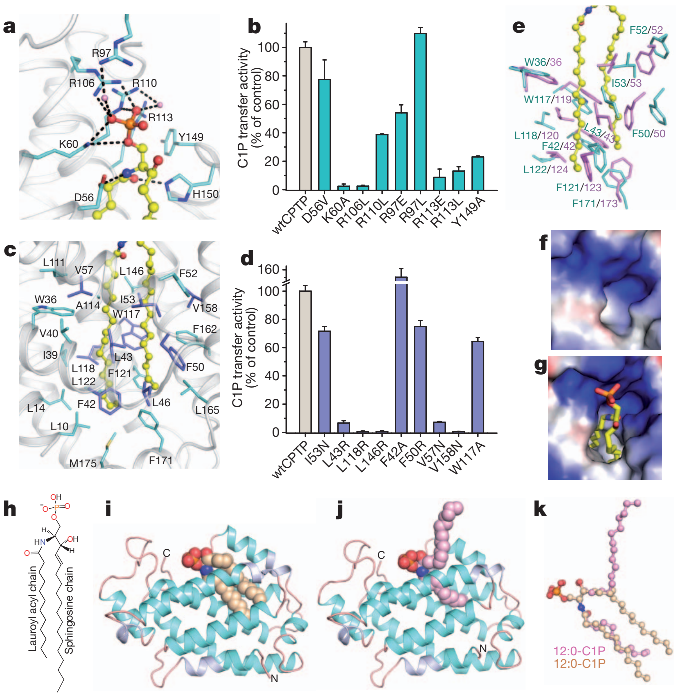


Figure 2 | CPTP conformation and functional recognition of C1P. **a**, CPTP lipid headgroup recognition centre residue interaction with phosphate and amide groups of bound 16:0-C1P (ball-and-stick). Hydrogen bonds shown as dashed lines. CPTP Cα backbone is light grey; side chains, cyan; and oxygen and nitrogen, red and blue, respectively. Water molecules are pink spheres. **b**, C1P transfer by CPTP point mutants of phosphate headgroup recognition cavity (cyan) or wild-type CPTP (wtCPTP, grey). **c**, Non-polar residues forming hydrophobic pocket that accommodates 16:0-C1P sphingosine and acyl chains. **d**, C1P transfer by CPTP point mutants (violet) of the hydrophobic pocket. wtCPTP (grey); Side-chains shown in panel **c**. Data in **b** and **d** represent the mean ± s.d. of three independent experiments. **e**, Conformational changes in hydrophobic pocket upon 16:0-C1P binding. Side-chains of apo-CPTP (lavender; stick) and human CPTP (cyan; stick) with bound 16:0-C1P (yellow; ball-and-stick). **f**, **g**, Surface electrostatics of hydrophobic pocket opening at lipid headgroup recognition sites in apo-CPTP (**f**) and 16:0-C1P–CPTP complex (**g**). **h**, 12:0-C1P chemical formula. **i**, **j**, Crystal structures of CPTP (ribbon) with bound 12:0-C1P in sphingosine-in (**i**; beige) and sphingosine-out (**j**; pink) conformations. **k**, Superposition of bound 12:0-C1P in sphingosine-in and sphingosine-out conformations.

key residues of the phosphate recognition site appear to be conserved in eukaryotes (Supplementary Fig. 2a).

The hydrophobic pocket is lined by ~25 nonpolar residues, mainly Phe, Leu, Val and Ile (Fig. 2c and Supplementary Fig. 2b) that prevent water entry while ensheathing the ceramide aliphatic chains. Mutation of L43, L118 or L146 to positively-charged Arg, or V57 or V158 to high polarity Asn compromises hydrophobic pocket functionality and strongly diminishes C1P transfer (Fig. 2d). More conservative mutation (for example, W117A) only moderately reduces C1P transfer, whereas F42A near the pocket bottom stimulates C1P transfer. Mutation near the entry portal (I53N) or in the flexible α1–2 loop (F50R) is well tolerated (75–80% active) (Fig. 2d). Ceramide entry is oriented by hydrogen bonding of the lipid amide oxygen and nitrogen with H150 and D56, respectively. Hydrogen bond disruption between lipid amide nitrogen and D56 (D56V) moderately slows C1P transfer, but H150 mutation (H150L) abolishes activity. Superpositioning of apo- and 16:0-C1P–CPTP structures (root mean squared deviation (r.m.s.d.) 1.4 Å) shows K60, R106 and R110 nearly identically positioned in the positively-charged surface cavities. Yet, large conformational differences exist for I53, W36, W119 and F52 (Fig. 2e) due to closer packing of certain α-helices in apo-CPTP (Supplementary Fig. 2d, e). Many Leu and Phe are repositioned, reducing the solvent accessible volume (40 $Å^3$) (Supplementary Table 4) and effectively collapsing the hydrophobic pocket (Fig. 2f, g) compared to the 16:0-C1P–CPTP complex (364 $Å^3$).

CPTP adaptability for different C1P species is reflected in structures of CPTP complexed with C1P containing differing-length acyl chains (Figs 2h–k, 3a–f and Supplementary Fig. 4c–f). Two lipid-binding conformational modes are apparent. In 'sphingosine-in' mode, both ceramide chains occupy the hydrophobic pocket, whereas only the acyl chain occupies the pocket in the 'sphingosine-out' mode. For the 12:0-C1P–CPTP complex (Fig. 2h–j), both binding modes occurred in the same asymmetric unit (Fig. 2i, j and Supplementary Figs 3c, d and 4a, b) enabling comparison (Fig. 2k) under the closest possible conditions. The lipid phosphate headgroups and amide groups bind exactly as in the 16:0-C1P–CPTP complex (Supplementary Figs 2b and 4a, b). In sphingosine-out mode (Fig. 2j and Supplementary Fig. 3d), a bend in sphingosine at C6 is stabilized by hydrophobic interactions with V153, V154 and the D56 Cβ atom, enabling outward projection. Sphingosine cross-bridging interactions with F50, I149, A157 and V153 of neighbouring, symmetry-related CPTP stabilize further (Supplementary Fig. 4b). Solvent accessible pocket volumes reflect the altered sphingosine location, that is, 261 $Å^3$ for sphingosine-out versus 329 $Å^3$ for sphingosine-in (Supplementary Table 4). In 18:1-C1P–CPTP complex (Fig. 3a, b), the *cis*-double bond kink in the acyl chain increases separation from the sphingosine chain, maximally expanding the pocket (stereo view; Supplementary Fig. 4c, d) although leaving the overall chain length in the pocket similar to 16:0-C1P. Accordingly, the solvent accessible volume of the hydrophobic pocket of 18:1-C1P–CPTP is larger (387 $Å^3$) (Supplementary Table 4) than in 16:0-C1P–CPTP where slightly closer packing by the saturated acyl chain decreases the solvent accessible volume (364 $Å^3$). Shortening the acyl chain length reduces solvent accessible pocket volumes to 104 $Å^3$ for 8:0 and 263 $Å^3$ for 2:0 (Supplementary Table 4). Structures for 2:0-C1P–CPTP (sphingosine-in) and 8:0-C1P–CPTP (sphingosine-out) are detailed in Supplementary Figs 3a, b and 4e, f and Supplementary Discussion.

The functional consequences of CPTP hydrophobic pocket structural adaptability become clear upon transfer analyses. Pocket expansion accommodates ceramide aliphatic chains in 'molecular ruler'-like fashion with CPTP adaptability limits optimized for 16:0- or 18:1-C1P species which are particularly effective competitors at slowing the AV-C1P transfer rate (Fig. 3g), consistent with maximal pocket expansion and optimal fit (Supplementary Table 4). It is noteworthy that C1P containing long lignoceryl (24:0) acyl chains are not very effective competitors, suggesting poor accommodation in the hydrophobic pocket

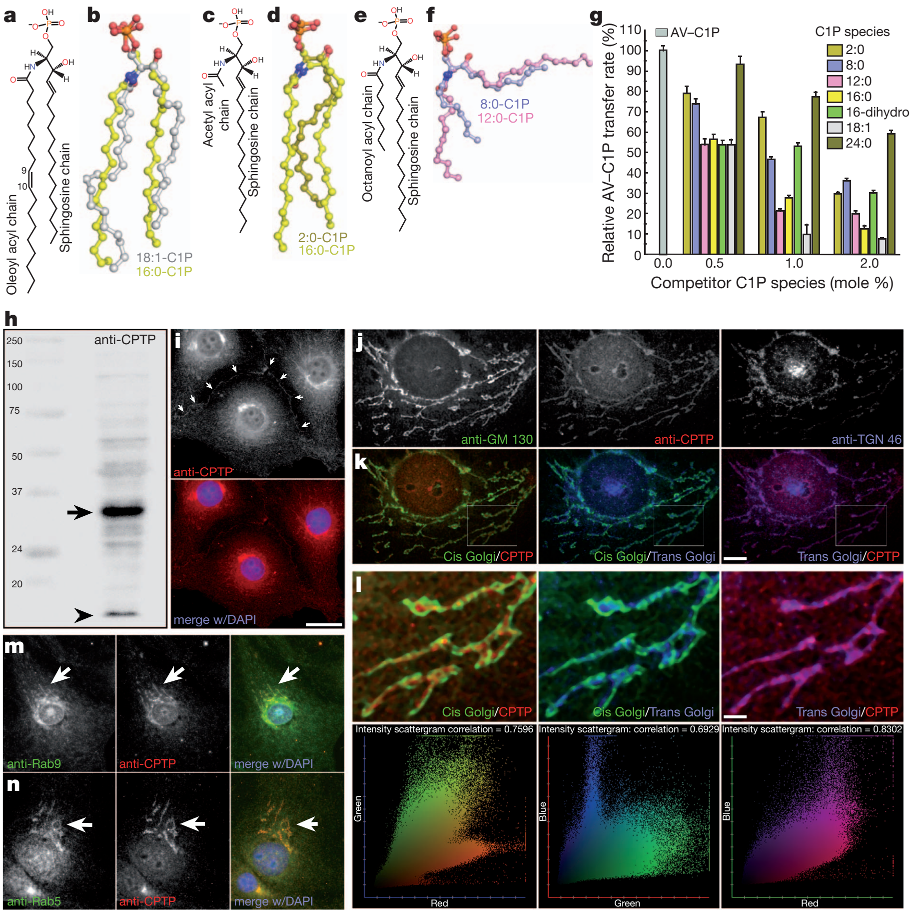


Figure 3 | CPTP accommodation and adaptation for C1P species, functional assessment of CPTP and intracellular localization. **a**, 18:1-C1P chemical formula. **b**, Superposition of bound 18:1-C1P (silver) and 16:0-C1P (yellow) lipid chains in their sphingosine-in CPTP complexes. **c**, 2:0-C1P chemical formula. **d**, Superposition of bound 2:0-C1P (olive) and 16:0-C1P (yellow) lipid chains in their sphingosine-in CPTP complexes. **e**, 8:0-C1P chemical formula. **f**, Superposition of bound 8:0-C1P (light blue) and 12:0-C1P (pink) lipid chains in their sphingosine-out CPTP complexes. **g**, Competition effects by C1P species containing different acyl chains on the CPTP-mediated initial transfer rate of fluorescent C1P. Data represent the mean ± s.d. of three independent experiments. **h–n**, CPTP localization to trans-Golgi and endosomes *in vivo*. **h**, SDS–polyacrylamide gel electrophoresis (SDS–PAGE) or immunoblot of BSC1 whole cell lysate, (anti-CPTP label), reveals a single immunoreactive band (arrow). Dye front position indicated by arrowhead. **i**, Anti-CPTP decorates perinuclear membrane stacks, nuclear membrane and plasma membrane (arrows). **j–l**, Perinuclear membrane stacks detected by anti-CPTP are trans-Golgi. **j**, BSC1 cell Golgi stacks labelled with anti-GM130 (cis-Golgi marker), anti-CPTP and anti-TGN 46 (trans-Golgi marker). **k**, Pseudo-colour overlay of frames from **j**. **l**, High magnification insets from **k** show CPTP/TGN-46 co-localization and intensity scattergram analyses with measured correlation coefficients[29]. **m–n**, Anti-CPTP co-localization with late (**m**, anti-Rab9 co-labelling) and early (**n**, anti-Rab5 co-labelling) endosomes. Wide-field fluorescence microscopy, scale bars, 10 μm (**i**), 2.5 μm (**k**), 1.0 μm (**l**).

because of adaptation limitations. Also, 16:0-C1P with dihydrosphingosine base competes less effectively than 16:0-C1P with naturally prevalent sphingosine base.

Structure determination of the di12:0-PA–CPTP complex elucidated the molecular basis of phosphatidic acid non-transfer (Supplementary Fig. 5a–h and Supplementary Table 1). Phosphatidic acid occupies the same binding site and its phosphate group interacts with the same positively charged residues as C1P (Supplementary Fig. 5b–d). Yet, K60 hydrogen bonding is single rather than bifurcated, and the lack of the acyl-amide moiety results in no hydrogen bonding with D56, distorting the position of the phosphate headgroup and both lipid chains and loosening phosphatidic acid binding. The distorted interaction mitigates phosphatidic acid transfer by CPTP (Supplementary Fig. 5e–h and Supplementary Discussion).

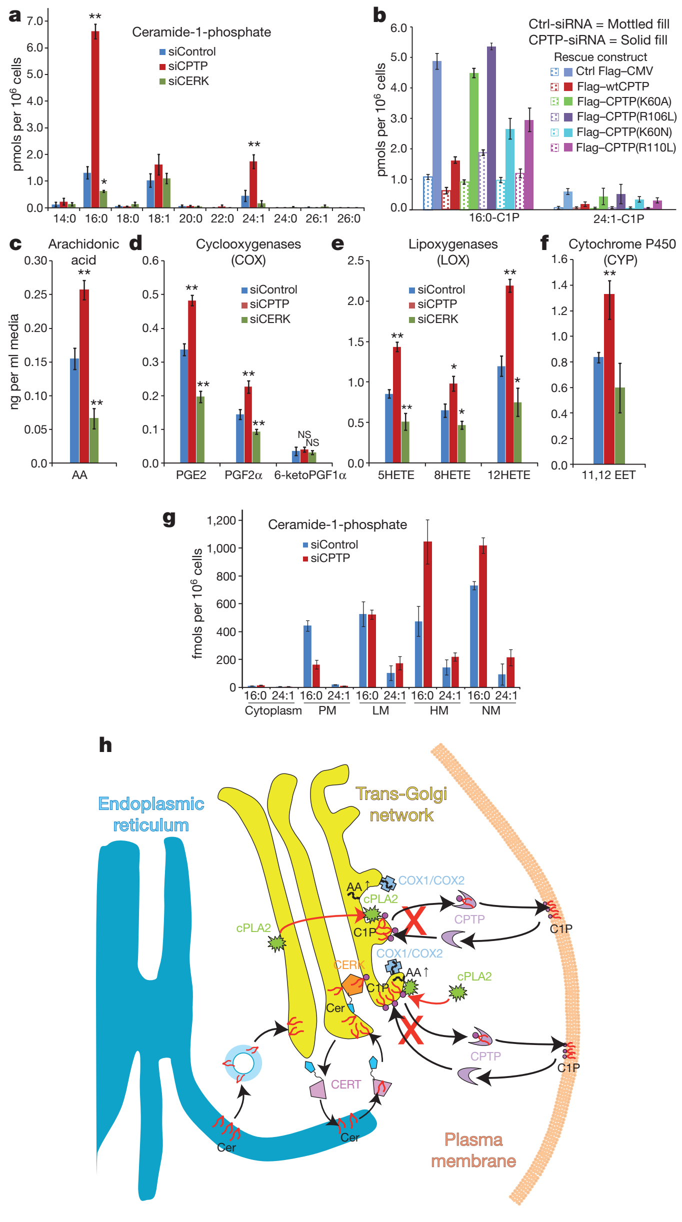
CPTP architecture not only represents a new motif for specific binding of phosphosphingolipids, but is previously unknown for any phosphate-modified biomolecule[15–17]. In CPTP, the fixed cationic residues of the phosphate recognition site undergo minimal conformational change upon C1P binding. B-factor distribution analyses show the regions between α1–α2 and α5–α6 are most flexible (Supplementary Fig. 6a, b), consistent with a cleft-like gating mechanism facilitating ceramide chain entry or exit. The conserved lipid orientation in the pocket, with the nonpolar acyl chain always inside regardless of sphingosine being in or out, supports a concerted mechanism of action in which the acyl chain enters first and leaves last during membrane interaction (Supplementary Fig. 6c, d and Supplementary Discussion).

The conformational adaptability of the inherently flexible, single-cavity, hydrophobic pocket of CPTP contrasts with lipid cavities in fatty acid binding proteins which use β-barrels/β-cups to generate a large, solvent-filled binding site that remains conformationally fixed whether or not occupied by fatty acid[18]. A single, fixed, lipid binding cavity also is characteristic of START lipid binding domains in PC transfer protein[19] and CERT (ref. 20), which uses an α/β fold built around an incomplete U-shaped β-barrel to bind ceramide[21] (Supplementary Fig. 7 and Supplementary Discussion).

In the human genome, the differing origins of CPTP and GLTP are clear. CPTP (214 amino acids) is encoded by a three-exon transcript originating from *GLTPD1* on chromosome 1 (locus 1p36.33). GLTP (209 amino acids) is encoded by a five-exon transcript originating from *GLTP* on chromosome 12 (locus 12q24.11)[22]. The shared folding topology encoded by *GLTPD1* and *GLTP*, despite only limited sequence homology (Supplementary Fig. 8a–e) and different lipid specificity, provides a striking example of evolutionary convergence and emphasizes the structural premium placed by eukaryotes on conservation of this fold[23–25]. The related architectures of CPTP and GLTP, but with naturally evolved and remarkably different lipid headgroup specificity (Supplementary Discussion), suggest that the term 'sphingolipid transfer protein (SLTP) superfamily' might better reflect the existence of the two major subfamilies: CPTP, with selectivity for ceramide-linked phosphates; and GLTP, with selectivity for ceramide-linked sugars.

In cells, CPTP tracked by monospecific antibody or fluorescent epitope tag enhanced green fluorescent protein (EGFP) is localized in the cytosol but also associates with perinuclear membranes (for example, Golgi/trans-Golgi network (TGN)/endosomes), nuclei and plasma membranes (Fig. 3i–n and Supplementary Fig. 9). No localization to mitochondria, lysosomes or the endoplasmic reticulum is detected. CPTP co-localization with TGN-46 verified interaction with the TGN, a site where ceramide kinase (CERK) generates C1P (refs 3, 26–28) and led us to propose a C1P regulatory/sensing role for CPTP during CERK-mediated metabolic/signalling events. Short interfering

RNA (siRNA)-induced CPTP downregulation (~90%; Supplementary Fig. 10a) elevated both 16:0-C1P and 24:1-C1P (~4-fold) (Fig. 4a) and fragmented the Golgi cisternal stacks (Supplementary Fig. 11). RNAi-induced C1P changes were partially rescued with moderately-active R110L and K60N, but not with inactive K60A and R106L mutants (Fig. 4b). CPTP overexpression in the absence of RNAi decreased 16:0-C1P and 24:1-C1P. K60A or R106L overexpression had the opposite effect (Supplementary Fig. 12) and fragmented the Golgi cisternal stacks (not shown) consistent with a dominant-negative effect. CPTP depletion measurably decreased sphingosine and S1P (Supplementary Fig. 10b, c), 14:0-, 22:0-, 24:1- and 24:0-sphingomyelin, 16:0-monohexosylceramide, and 24:0-Cer, but modestly increased 24:1-Cer (Supplementary Fig. 10d–f and Supplementary Discussion). Subcellular fractionation showed increased 16:0-C1P levels in 'heavy' membranes (TGN/endosome-enriched) and nuclear fractions, and decreased levels in plasma membranes without affecting levels in 'light' membranes (cis-Golgi/endoplasmic reticulum (ER)-enriched). These CPTP-depletion induced effects were rescued by ectopic wild-type CPTP expression (Fig. 4g, Supplementary Fig. 13c and Supplementary Discussion).



Because CERK-generated C1P induces Group IVA cPLA$_2\alpha$ activity which releases arachidonic acid used to produce pro-inflammatory eicosanoids[26,27], CPTP involvement was assessed. In siRNA-induced CPTP-depleted cells, arachidonic acid increased (Fig. 4c) consistent with C1P accumulation at the Golgi/TGN that activates cPLA$_2\alpha$ (ref. 26). Also elevated were major arachidonic acid metabolites generated by cyclooxygenase (COX), lipoxygenase (LOX) and cytochrome P450 (CYP) pathways, that is, PGE2, PGF2α, 6-keto-PGF1α (COX, Fig. 4d); 5HETE, 8HETE, 12HETE (LOX, Fig. 4e); 11,12 EET (CYP, Fig. 4f). By contrast, arachidonic acid and eicosanoid levels decreased upon overexpression of wild-type CPTP but not the K60A or R106L mutant (Supplementary Fig. 14). Parallel siRNA-induced downregulation of CERK (Supplementary Fig. 10), the only established producer of C1P in mammals, decreased 16:0-C1P (Fig. 4a), arachidonic acid (Fig. 4c), and eicosanoids (Fig. 4d–f) elevated by CPTP depletion.

Figure 4h depicts a model showing how CPTP could regulate pro-inflammatory eicosanoid generation. In mammals, the only established pathway for C1P production is through phosphorylation of ceramide by CERK at the cytoplasmic surface of the TGN[3,28]. CERK also contains nuclear localization/export signals and traffics to the plasma membrane by microtubule-driven vesicles in response to hyperosmotic shock[28]. To produce C1P, CERK uses ceramide delivered from its endoplasmic reticulum synthetic site to the Golgi by either CERT (ref. 27) or possibly by vesicular trafficking[10]. C1P elevation by CERK is known to activate soluble cPLA$_2\alpha$ by enhancing translocation to the TGN[3,26], where cPLA$_2\alpha$ action releases arachidonic acid needed by eicosanoid producers such as COX-1 or COX-2. siRNA-induced CPTP depletion elevates C1P in the Golgi complex and nucleus, but lowers C1P plasma membrane levels. We propose that CPTP prevents excess C1P accumulation after production by CERK, thereby regulating cPLA$_2\alpha$ action,

Figure 4 | CPTP siRNA depletion/rescue effects on cellular C1P levels, arachidonic acid generation and eicosanoid release and model of CPTP cell biological function. a, Intracellular C1P levels of siRNA-treated and control A549 cells. Acyl composition of C1P species with sphingosine base chains ($d_{18:1}$) shown on the *x* axis. **b,** Rescue effect by partially active K60N and R110L CPTP mutants, but not inactive K60A and R106L CPTP mutants with siRNA targeted to endogenous the CPTP 3′ UTR. **c–f,** Arachidonic acid (AA) and eicosanoids (COX, LOX and CYP pathways) secreted into media by siRNA-treated and control A549 cells (blue, siControl; red, siCPTP; green siCERK). Data for **b** and **d–g** represent averages of 6 experiments (2 procedures) by Student's *t*-test (*$P < 0.05$, **$P < 0.01$; NS, not significant). 6keto-PGF1α, the primary prostacyclin metabolite, is nearly unaffected by siCPTP or siCERK suggesting existence of an arachidonic acid pool derived independently of C1P-activated cPLA$_2\alpha$. **g,** Changes in C1P levels at various subcellular locations by CPTP-siRNA. Shown on the *x* axis are the C1P species containing sphingosine base chains ($d_{18:1}$) and either 16:0 or 24:1 acyl chains. PM, plasma membranes; LM, light membranes (ER/cis-Golgi enriched); HM, heavy membranes (trans-Golgi/endosome/mitochondria enriched); NM, nuclear membranes. Data represent the mean ± s.d. of three independent experiments. **h,** Model for CPTP regulation of eicosanoid production. C1P is synthesized by ceramide kinase (CERK) which concentrates in the trans-Golgi network vicinity via its pleckstrin homology (PH) domain during stimulation. To produce C1P, CERK uses ceramide transported from the ER by ceramide transfer protein (CERT) which also contains a targeting PH domain. After synthesis, C1P is transported to subcellular destinations by CPTP and possibly by vesicular trafficking. RNAi knockdown of CPTP (shown by the red x) leads to accumulation and elevation of C1P at the Golgi complex, a condition that activates soluble cytosolic phospholipase A$_2\alpha$ (cPLA$_2\alpha$), releasing arachidonic acid needed for generation of downstream, pro-inflammatory eicosanoids. CPTP overexpression has the opposite effect on lipid levels. cPLA$_2\alpha$ activation occurs by translocation from the cytoplasm and/or cis-Golgi (red arrows) through enhanced anchoring to C1P generated by CERK in the TGN. COX-1 and inducible COX-2 which use arachidonic acid to produce pro-inflammatory prostaglandins, also concentrate in the TGN vicinity during stimulation. For clarity, other eicosanoid generator pathways, that is, LOX (for example, cytoplasmic 5-lipoxygenase) and CYP (ER-associated cytochrome P450), are not depicted. Also not depicted is Golgi cisternal stack fragmentation induced by 24 h of CPTP RNAi.

diminishing arachidonic acid release and downstream generation of pro-inflammatory eicosanoids. One destination for CPTP cargo is the plasma membrane. Models involving CPTP in catabolic C1P generation by sphingomyelinase D are less plausible because mammalian cells lack this enzyme[3] (Supplementary Discussion).

Previously, the only identified mechanism for regulating CERK-mediated production of C1P was by control of ceramide availability through ceramide transfer protein[27]. It is noteworthy that siRNA-induced CPTP depletion yields the highest increase in endogenous C1P reported to date, mostly as 16:0-C1P, and dramatically alters Golgi cisternal stack morphology indicating CPTP-mediated transport is essential for maintaining proper Golgi organization by safeguarding localized C1P levels. The ensuing stimulation in eicosanoid production triggered by elevated C1P in the Golgi complex potentially implicates CPTP in, as of yet unidentified, disease states associated with inflammation.

METHODS SUMMARY

CPTP was cloned and expressed in BL21 (DE3) Star cells using pET-SUMO vector. Lipid intervesicular transfer assays were performed by Förster resonance energy transfer. X-ray diffraction data were collected on crystals of mouse apo-CPTP and of human CPTP with bound C1P containing acyl chains of differing length, for example, 2:0, 8:0, 12:0, 16:0 and 18:1, as well as di-12:0 PA. For phasing, single-wavelength anomalous dispersion data were collected for Se-Met-labelled, 8:0-C1P–CPTP crystal complex at Se peak wavelength (Supplementary Table 2) enabling other structures to be solved by molecular replacement. Data collection, processing, structure solution and refinement are described in the Methods. Protein docking with membranes was performed using the Orientation of Proteins with Membranes modelling. Epifluorescence microscopy images of fixed BSC-1 cells were captured by labelling with anti-CPTP antibody and anti-TGN-46, anti-GM130, anti-p230, anti-Rab5 or anti-Rab9 followed by secondary antibodies coupled to Alexa Fluor 488, Alexa Fluor 594, or Alexa Fluor 660 and counterstaining with DAPI (4′,6-diamidino-2-phenylindole). Reverse transcription PCR and quantitative PCR were used to evaluate endogenous *CPTP* mRNA tissue levels and siRNA-downregulated *CPTP* and *CERK* transcript levels. Sphingolipids were analysed by electrospray ionization tandem mass spectrometry (ESI-MS/MS) after separation by HPLC. Eicosanoids were analysed as detailed by the Lipid Maps Consortium.

Full Methods and any associated references are available in the online version of the paper.

Received 13 July 2012; accepted 22 May 2013.
Published online 17 July 2013.

1. Hannun, Y. A. & Obeid, L. M. Principles of bioactive lipid signalling: lessons from sphingolipids. *Nature Rev. Mol. Cell Biol.* **9,** 139–150 (2008).
2. Chalfant, C. E. & Spiegel, S. Sphingosine 1-phosphate and ceramide 1-phosphate: expanding roles in cell signaling. *J. Cell Sci.* **118,** 4605–4612 (2005).
3. Lamour, N. F. & Chalfant, C. E. Ceramide kinase and the ceramide-1-phosphate/cPLA2α interaction as a therapeutic target. *Curr. Drug Targets* **9,** 674–682 (2008).
4. Gangoiti, P. *et al.* Control of metabolism and signaling of simple bioactive sphingolipids: Implications in disease. *Prog. Lipid Res.* **49,** 316–334 (2010).
5. Maceyka, M., Harikumar, K. B., Milstien, S. & Spiegel, S. Sphingosine-1-phosphate signaling and its role in disease. *Trends Cell Biol.* **22,** 50–60 (2012).
6. Wymann, M. P. & Schneiter, R. Lipid signalling in disease. *Nature Rev. Mol. Cell Biol.* **9,** 162–176 (2008).
7. Leslie, C. C., Gangelhoff, T. A. & Gelb, M. H. Localization and function of cytosolic phospholipase A2α at the Golgi. *Biochimie* **92,** 620–626 (2010).
8. Bechler, M. E., de Figueiredo, P. & Brown, W. J. A. PLA1–2 punch regulates the Golgi complex. *Trends Cell Biol.* **22,** 116–124 (2012).
9. Harizi, H., Corcuff, J. B. & Gualde, N. Arachidonic-acid-derived eicosanoids: roles in biology and immunopathology. *Trends Mol. Med.* **14,** 461–469 (2008).
10. Boath, A. *et al.* Regulation and traffic of ceramide 1-phosphate produced by ceramide kinase: comparative analysis to glucosylceramide and sphingomyelin. *J. Biol. Chem.* **283,** 8517–8526 (2008).
11. Lev, S. Non-vesicular lipid transport by lipid-transfer proteins and beyond. *Nature Rev. Mol. Cell Biol.* **11,** 739–750 (2010).
12. Prinz, W. A. Lipid trafficking sans vesicles: where, why, how? *Cell* **143,** 870–874 (2010).
13. Malinina, L., Malakhova, M. L., Teplov, A., Brown, R. E. & Patel, D. J. Structural basis for glycosphingolipid transfer specificity. *Nature* **430,** 1048–1053 (2004).
14. Malakhova, M. L. *et al.* Point mutational analysis of the liganding site in human glycolipid transfer protein. Functionality of the complex. *J. Biol. Chem.* **280,** 26312–26320 (2005).
15. Hirsch, A. K., Fischer, F. R. & Diederich, F. Phosphate recognition in structural biology. *Angew. Chem. Int. Ed. Engl.* **46,** 338–352 (2007).
16. Bourquin, F., Riezman, H., Capitani, G. & Grutter, M. G. Structure and function of sphingosine-1-phosphate lyase, a key enzyme of sphingolipid metabolism. *Structure* **18,** 1054–1065 (2010).
17. Berna, A. *et al.* For whom the bell tolls? DING proteins in health and disease. *Cell. Mol Life Sci.* **66,** 2205–2218 (2009).
18. Furuhashi, M. & Hotamisligil, G. S. Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets. *Nature Rev. Drug Discov.* **7,** 489–503 (2008).
19. Roderick, S. L. *et al.* Structure of human phosphatidylcholine transfer protein in complex with its ligand. *Nature Struct. Biol.* **9,** 507–511 (2002).
20. Hanada, K. *et al.* Molecular machinery for non-vesicular trafficking of ceramide. *Nature* **426,** 803–809 (2003).
21. Kudo, N. *et al.* Structural basis for specific lipid recognition by CERT responsible for nonvesicular trafficking of ceramide. *Proc. Natl Acad. Sci. USA* **105,** 488–493 (2008).
22. Zou, X. *et al.* Human glycolipid transfer protein (*GLTP*) genes: organization, transcriptional status and evolution. *BMC Genomics* **9,** 72 (2008).
23. Orengo, C. A. & Thornton, J. M. Protein families and their evolution - a structural perspective. *Annu. Rev. Biochem.* **74,** 867–900 (2005).
24. Galperin, M. Y. & Koonin, E. V. Divergence and convergence in enzyme evolution. *J. Biol. Chem.* **287,** 21–28 (2012).
25. Liberles, D. A. *et al.* The interface of protein structure, protein biophysics, and molecular evolution. *Protein Sci.* **21,** 769–785 (2012).
26. Lamour, N. F. *et al.* Ceramide 1-phosphate is required for the translocation of group IVA cytosolic phospholipase A2 and prostaglandin synthesis. *J. Biol. Chem.* **284,** 26897–26907 (2009).
27. Lamour, N. F. *et al.* Ceramide kinase uses ceramide provided by ceramide transport protein: localization to organelles of eicosanoid synthesis. *J. Lipid Res.* **48,** 1293–1304 (2007).
28. Bornancin, F. Ceramide kinase: the first decade. *Cell. Signal.* **23,** 999–1008 (2011).
29. Durcan, T. M. *et al.* Tektin 2 is required for central spindle microtubule organization and the completion of cytokinesis. *J. Cell Biol.* **181,** 595–603 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements This research was supported by NIH/NCI CA121493 (D.J.P. & R.E.B.), NIH/NIGMS GM45928 (R.E.B.), NIH/NIGMS GM072754 (E.H.H.), NIH/CA154314 (C.E.C.), VA Merit Award (C.E.C.), VA Research Career Scientist Award (C.E.C.), VA Career Devel. Award (D.S.W.), NRS-T32/NIGMS 008695 (D.S.W.), Spanish Ministerio de Ciencia e Innovacion BFU2010-17711 (L.M.), Russian Foundation for Basic Research #12-04-00168 (J.G.M.), Hormel Foundation. (R.E.B.), Abby Rockefeller Mauze Trust (D.J.P.) and Maloris Foundation (D.J.P.). We thank H. Pike for expressing and purifying protein used for transfer activity analyses, K. Karanjeet for preparing cells for confocal and epifluorescence microscopy, and the staff of X-29 beamline at the National Synchrotron Light Source and ID-24-C/E beamlines at the Advanced Photon Source for help.

Author Contributions D.K.S. carried out all structural analyses and provided evidence for C1P binding by CPTP, generated all CPTP point mutants and wrote text. R.K.K did transfer analyses of wild-type CPTP and CPTP point mutants and wrote text. D.S.W. conducted siRNA CPTP knockdown, rescue, and all lipid analyses and wrote text. Xi.Zo. cloned wild-type CPTP and did PCR analyses of CPTP transcript distribution in human tissues. Xi.Zh. did CPTP transfer rate analyses. S.K.M. prepared CPTP RNAi constructs for microscopy and CPTP overexpression constructs for lipidomics analyses. J.G.M. synthesized all fluorescent lipids. L.M. contributed to structural data interpretation. E.H.H. did fluorescence microscopy of CPTP localization in fixed and living cells and finalized the write-up. C.E.C. directed siRNA knockdown, rescue and related lipidomics analyses and finalized the write-up. D.J.P. directed CPTP structural analyses and finalized the write-up. R.E.B. directed functional analyses after the initial CPTP discovery in his laboratory, finalized the write-up and coordinated and integrated all section write-ups.

Author Information Atomic coordinates and structure factors for human CPTP crystal complexes with various lipids and mouse apo-CPTP have been deposited in the Protein Data Bank. Accession codes are: 2:0-C1P–CPTP (4K80), 8:0-C1P–CPTP (4KF6), 12:0-C1P–CPTP (4K85), 16:0-C1P–CPTP (4K84), 18:1-C1P–CPTP (4K8N), di12:0-PA–CPTP (4KBS) and mouse apo-CPTP (4KBR). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.E.B. (reb@umn.edu), E.H.H. (ehinchcliffe@hi.umn.edu), C.E.C. (cechalfant@vcu.edu) or D.J.P. (pateld@mskcc.org).

METHODS

Protein expression and purification. *GLTPD1* ORFs encoding human and mouse (GenBank JN542538 & NP_077792.2) CPTP were cloned in pET-SUMO vector (Invitrogen) and expressed in BL21 (DE3) Star cells (Invitrogen). Soluble CPTP tagged N-terminally with His_6–SUMO was affinity-purified by Ni-NTA chromatography followed by ubiquitin-like protein 1 (Ulp1) SUMO protease digestion overnight at 4 °C to release the His_6–SUMO tag. Affinity repurification by Ni-NTA chromatography was followed by FPLC gel filtration chromatography. L-selenomethionine (Se-Met)-labelled protein for *ab initio* phasing was produced by feedback inhibition of the methionine synthesis pathway. Mutants were constructed by QuikChange Site-Directed Mutagenesis (Stratagene) and verified by sequencing.

CPTP lipid transfer activity involving membrane vesicles. Intermembrane lipid transfer by CPTP was measured in real time by Förster resonance energy transfer (FRET) between donor POPC vesicles, containing 1 mole % AV-lipid (acyl chain omega-labelled with anthrylvinyl fluorophore, that is, (11*E*)-12-(9-anthryl)-11-dodecenoyl), and 1.5 mole % 1-acyl-2-[9-(3-perylenoyl)-nonanoyl]-3-*sn*-glycero-3-phosphocholine (Per-PC) and POPC acceptor vesicles at tenfold excess. In competition assays, donor vesicles also contained competitor lipids at 0.5, 1.0 and 2.0 mole % (ref. 30). Briefly, CPTP addition produces an exponential increase in AV emission intensity (425 nm) as the protein transports AV-C1P away from the donor vesicles (creating separation from the 'nontransferable' Per-PC) and delivers to the POPC acceptor vesicles present in tenfold excess. The time-dependent increase in 425 nm emission relative to signal in the absence of CPTP reflects lipid transfer kinetics. In the absence of acceptor vesicles, no transfer is observed. The initial lipid transfer rate, ν_o, is obtained by nonlinear regression analyses (see Supplementary Methods).

Crystallization and structure determination. Crystallization hits from initial screens were optimized by the hanging drop vapour diffusion method and systematically varying pH and individual component concentrations (Supplementary Table 3). For data collection, crystals were flash frozen (100 K) in reservoir solutions containing 20% (v/v) ethylene glycol. Diffraction data sets were collected on 24-ID-C and 24-ID-E beamlines at the Advanced Photon Source (APS) and X29 beamline at Brookhaven National Laboratory. All crystals belonged to different crystal forms. For phasing, single-wavelength anomalous dispersion (SAD) data were collected for Se-Met-labelled, 8:0-C1P–CPTP crystal complex at Se peak wavelength (Supplementary Table 2; see Supplementary Methods). Use of Se-CPTP structure enabled other structures to be solved by molecular replacement (Supplementary Methods). Statistics for data collection, refinement and SAD phasing are provided in Supplementary Tables 1 and 2.

Epifluorescence microscopy analyses. BSC-1 cells on coverslips were fixed in −20 °C methanol and labelled with anti-CPTP (Santa Cruz Biotechnology, sc247014), and Golgi markers anti-TGN46, anti-GM 130 and anti-p230, or endosome markers anti-Rab5 and anti-Rab9 (Cell Signaling) followed by secondary antibodies coupled to Alexa Fluor 488, Alexa Fluor 594 or Alexa Fluor 660. Cells were counter-stained with DAPI, mounted in 10% PBS, 90% glycerol, imaged using a Leica DM RXA2 microscope with a ×63 1.4 NA APO C objective, a Hamamatsu ORCA ER CCD camera, and Simple PCI software and analysed as intensity scattergrams with measured correlation coefficients[31]. Time-lapse images of living cells expressing EGFP–CPTP were captured with a Leica DM RXA2 microscope stand equipped with a Yokagawa CSU-10 spinning disk confocal head and using illumination from a Coherent 488 nm 200 mW 'Sapphire' continuous wave optically pumped solid-state laser and 3I Slidebook software[32] (see Supplementary Methods for more details).

Immunoblot analysis. BSC1 cells were grown to semi-confluence, collected by manual scraping, pelleted, and boiled in SDS–PAGE buffer. Proteins were separated on a 10% discontinuous SDS–PAGE gel, transferred to PDVF membrane and immuno-labelled[29]. The immunoreactive band was detected by chemiluminescence (Image Quant system, GE Healthcare).

siRNA-mediated CPTP downregulation and rescue in cells ectopically expressing wild-type and mutant CPTP constructs. Low passage A549 cells (5×10^5) were grown (10-cm plates) in appropriate medium under standard incubator conditions overnight. Cells were treated with siRNA (Dharmacon) against CPTP (*GLTPD1*) or CERK as well as non-targeting siRNA sequence for control per manufacturer's protocol and incubated for 48 h under standard incubator conditions. For rescue experiments, cells were transfected with either the empty pFLAG-CMV4 (Neo^r) plasmid or this vector containing wild-type CPTP, CPTP(K60A), CPTP(K60N), CPTP(R106L) or CPTP(R110L). Batch cultures of cells stably expressing the transfected constructs were obtained by selection for two weeks in regular medium containing G418 (genticin, 500 μg ml^{-1}) under standard incubator conditions. Following selection, cells (5×10^5) were transferred to 10-cm tissue culture plates and cultured overnight in regular media without G418 under standard incubator conditions. Cells then were treated with either control siRNA or a mixture of 4 siRNA constructs (Dharmacon) designed against the 3′ UTR of endogenous CPTP (*GLTPD1*) mRNA following standard manufacturer's protocol (Supplementary Methods). The 3′UTR was not included in the ectopically expressed constructs to ensure siRNA targeting only to endogenous CPTP. Cells were incubated 48 h in regular media without G418 under standard incubator conditions. Full serum media was replaced with media containing 2% serum 15 h before harvest.

RNA isolation, RT–PCR and quantitative PCR. To evaluate downregulation of CERK and CPTP, quantitative PCR was performed[26]. Briefly, total RNA was isolated using RNeasy kits (Qiagen). Total RNA (1 μg) was reverse transcribed using Superscript III reverse transcriptase (Invitrogen). The level of *CERK* transcript was monitored using quantitative PCR and TaqMan technology (Applied Biosystems) specific to CERK and CPTP with 18S rRNA as control. cDNA was amplified using an ABI 7900HT with premixed primer-probe sets and TaqMan Universal PCR master mix (Applied Biosystems).

Intracellular sphingolipid analyses. Cell lipids were harvested using an improved Bligh-Dyer protocol[29] (Supplementary Methods). Sphingolipids were separated by HPLC (Prominence HPLC system, Shimadzu) using a Kinetix-C18 column (50 × 2.1 mm, 2.6 μm; Phenomenex) and eluted using a linear gradient (solvent A, methanol:water:formic acid (58:41:1) in 5 mM ammonium formate; solvent B, methanol:formic acid (99:1) in 5 mM ammonium formate, 20–100% B (3.5 min) and at 100% B (4.5 min); flow rate of 0.4 ml min^{-1}, 60 °C). ESI-MS/MS (API 4000 QTRAP instrument; Applied Biosystems, MDS Sciex) was used to detect C1P (ref. 29), ceramide, sphingosine, S1P, sphingomyelin, and monohexosyl ceramide under positive ionization (see Supplementary Methods).

Eicosanoid analysis. Eicosanoids were analysed as detailed by the Lipid Maps Consortium[33,34]. Culture media (4 ml) from siRNA was combined with 10% methanol (400 μl) and glacial acetic acid (20 μl) before spiking with internal standard (100 μl) containing the following deuterated eicosanoids (100–pg–μl^{-1}, 10 ng total): (d_4) 6keto-PGF$_1\alpha$, (d_4) PGF$_2\alpha$, (d_4) PGE$_2$, (d_4) PGD$_2$, (d_8) 5-hydroxyeicosatetranoic acid (5HETE), (d_8) 15-hydroxyeicosatetranoic acid (15HETE), (d_8) 14,15 epoxyeicosatrienoic acid and (d_8) arachidonic acid. Samples and vial rinses (5% MeOH; 2 ml) were applied to Strata-X SPE columns (Phenomenex), previously washed with methanol (2 ml) and then dH_2O (2 ml). Eicosanoids eluted with isopropanol (2 ml), were dried *in vacuuo* and reconstituted in EtOH:dH_2O (50:50; 100 μl) before HPLC ESI-MS/MS analysis (see Supplementary Methods).

Subcellular fractionation. Subcellular fractionation was performed by multi-step centrifugation as detailed and characterized previously[27] with minor modifications (see Supplementary Methods). Fraction enrichment was validated by SDS–PAGE/western blotting (Supplementary Fig. 13b) using organelle markers for: nuclei (anti-lamin AC), trans-Golgi (anti-TGN46), ER (anti-protein disulphide isomerase (PDI)) and plasma membrane (anti-caveolin-1).

30. Samygina, V. R. *et al.* Enhanced selectivity for sulfatide by engineered human glycolipid transfer protein. *Structure* **19,** 1644–1654 (2011).
31. Manders, E. M. M., Stap, J., Brakenhoff, G. J., van Driel, R. & Aten, J. A. Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy. *J. Cell Sci.* **103,** 857–862 (1992).
32. Hornick, J. E. *et al.* Amphiastral mitotic spindle assembly in vertebrate cells lacking centrosomes. *Curr. Biol.* **21,** 598–605 (2011).
33. Wijesinghe, D. S. *et al.* Use of high performance liquid chromatography-electrospray ionization-tandem mass spectrometry for the analysis of ceramide-1-phosphate levels. *J. Lipid Res.* **51,** 641–651 (2010).
34. Blaho, V. A., Buczynski, M. W., Brown, C. R. & Dennis, E. A. Lipidomic analysis of dynamic eicosanoid responses during the induction and resolution of Lyme arthritis. *J. Biol. Chem.* **284,** 21599–21612 (2009).

LETTER

doi:10.1038/nature12313

The histone H4 lysine 16 acetyltransferase hMOF regulates the outcome of autophagy

Jens Füllgrabe[1], Melinda A. Lynch-Day[2], Nina Heldring[3], Wenbo Li[4], Robert B. Struijk[1], Qi Ma[4,5], Ola Hermanson[3], Michael G. Rosenfeld[4], Daniel J. Klionsky[2] & Bertrand Joseph[1]

Autophagy is an evolutionarily conserved catabolic process involved in several physiological and pathological processes[1,2]. Although primarily cytoprotective, autophagy can also contribute to cell death; it is thus important to understand what distinguishes the life or death decision in autophagic cells[3]. Here we report that induction of autophagy is coupled to reduction of histone H4 lysine 16 acetylation (H4K16ac) through downregulation of the histone acetyltransferase hMOF (also called KAT8 or MYST1), and demonstrate that this histone modification regulates the outcome of autophagy. At a genome-wide level, we find that H4K16 deacetylation is associated predominantly with the downregulation of autophagy-related genes. Antagonizing H4K16ac downregulation upon autophagy induction results in the promotion of cell death. Our findings establish that alteration in a specific histone post-translational modification during autophagy affects the transcriptional regulation of autophagy-related genes and initiates a regulatory feedback loop, which serves as a key determinant of survival versus death responses upon autophagy induction.

Autophagy is a catabolic process that results in the autophagosome-dependent lysosomal degradation of bulk cytoplasmic contents, abnormal protein aggregates and excess or damaged organelles[1,2]. This process involves a series of dynamic membrane-rearrangements mediated by a core set of autophagy-related (ATG) proteins[4]. Although autophagy is primarily a protective process for the cell, it can also play a role in cell death[3]; however, it is not clear what distinguishes the life or death decision[5]. Protein acetylation, in particular acetylation of ULK1 and ATG proteins, has emerged as a regulator of autophagy[6–9]. Accumulating evidence has established sirtuin 1 (SIRT1), a NAD^+-dependent deacetylase, as a player in this process[7,10]. However, SIRT1 is not always required for autophagy to occur. Indeed, although SIRT1 overexpression is sufficient to increase the basal level of autophagy and is required for starvation-induced autophagy, it is not necessary for rapamycin-induced autophagy[7,11]. SIRT1 has a wide range of non-histone targets, but lysine 16 on histone H4 (H4K16) is its primary histone target[12,13]. The histone acetyltransferase hMOF/KAT8/MYST1 is necessary and sufficient for the bulk of H4K16 acetylation and thereby antagonizes the enzymatic activity of SIRT1 (refs 13–15). Because SIRT1 has been linked to both autophagy and epigenetic chromatin changes, this encouraged us to investigate the role of covalent histone modifications in autophagy. As SIRT1 preferentially deacetylates H4K16ac[12], we considered that this histone modification could be altered upon autophagy induction.

We induced autophagy in mouse embryonic fibroblast (MEF) cells by amino-acid starvation and observed a decrease in acetylation of H4K16 (Fig. 1a). To investigate whether the observed effect on H4K16ac was linked to the role of SIRT1 during starvation-induced autophagy, or if the deacetylation of H4K16 is a general feature of the autophagic process, treatments with rapamycin or Torin 1, respectively allosteric and catalytic inhibitors of the kinase mechanistic target of rapamycin (MTOR), were used to induce SIRT1-independent autophagy (Supplementary Figs 2a and 3a). The global amount of H4K16ac was robustly reduced in MEF cells after those treatments (Fig. 1b and Supplementary Fig. 4a). Interestingly, in the *sirt1* null MEF cells, both rapamycin and Torin 1 treatments, but not amino-acid starvation, induced the downregulation of H4K16ac, confirming that SIRT1 is not required for the repression of this histone modification (Fig. 1c and Supplementary Figs 2b and 4b).

The downregulation of H4K16ac upon autophagy induction occurred in various human cancer cell types, namely U1810, HeLa and U2OS cells (Fig. 1d-e and Supplementary Fig. 4c, d) and was even found to occur in yeast (Supplementary Fig. 5). Rapamycin treatment did not affect total histone H4 amounts (Supplementary Fig. 6). The changes in H4K16 acetylation status were linked to the occurrence of autophagy as established by an increased lipidation of the autophagic marker LC3, resulting in an increased ratio of the lipidated form (LC3-II) to the unlipidated form (LC3-I), referred to later as LC3 conversion (Fig. 1a–d and Supplementary Fig. 3a, b). Similarly, in yeast this treatment resulted in increased lipidation of Atg8 (yeast homologue of LC3), and both cleavage and vacuolar localization of green fluorescent protein (GFP)-tagged Atg8 (Supplementary Fig. 7a–c).

Dynamic histone modifications play a pivotal role in cell regulatory events[16] and the H4K16 residue is of particular interest, as its acetylation influences higher-order chromatin structure[17] and plays an important

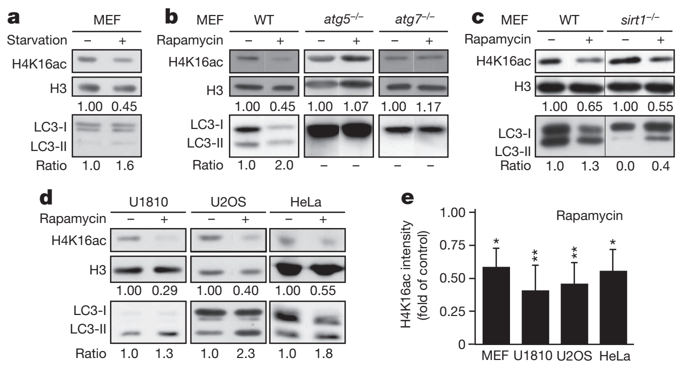


Figure 1 | Autophagy is associated with reduced acetylation of histone H4 lysine 16. **a**, Starvation (3 h)-induced autophagy results in a downregulation of H4K16ac in histone extracts of MEF cells. **b**, Upon rapamycin treatment (300 nM) LC3 conversion and downregulation of H4K16ac are observed in WT MEF cells but not in the autophagy-deficient $atg5^{-/-}$ and $atg7^{-/-}$ MEF cells. **c**, Rapamycin treatment increased the LC3-II/LC3-I ratio and promoted H4K16ac decrease in $sirt1^{-/-}$ and WT MEF cells. **d**, Rapamycin-induced autophagy led to downregulation of H4K16ac at 48 h in histone extracts of HeLa and U20S cells, and after 6 h in U1810 cells. **e**, Quantification of H4K16 acetylation by immunoblotting is depicted for rapamycin-treated cells. Data are expressed as mean ± s.e.m. (n = 3–5); $*P < 0.05$; $**P < 0.01$.

[1]Department of Oncology Pathology, Cancer Centrum Karolinska, Karolinska Institutet, Stockholm 17176, Sweden. [2]Life Sciences Institute and Departments of Molecular, Cellular and Developmental Biology and Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA. [3]Department of Neuroscience, Karolinska Institutet, Stockholm 17177, Sweden. [4]Howard Hughes Medical Institute, Department of Medicine, School of Medicine, University of California, San Diego, La Jolla, California 92093, USA. [5]Graduate Program in Bioinformatics and Systems Biology, University of California, San Diego, La Jolla, California 92093, USA.

role in transcription[18]. Chromatin immunoprecipitation targeting H4K16ac, followed by high-throughput sequencing (ChIP-seq) was performed to investigate the genome-wide occurrence of this histone mark in U1810 cells undergoing autophagy (Fig. 2a). H4K16ac ChIP-seq data analysis showed 3,422 called peaks in untreated U1810 cells, which subsequently showed reduced H4K16ac occupancy after 8 h rapamycin treatment. To gain insight into the role of this induced H4K16 deacetylation in the regulation of gene expression during autophagy, we performed a global run-on-sequencing (GRO-seq) assay[19,20] to generate a genome-wide view of the location, orientation and density of nascent transcripts engaged by RNA polymerases at high resolution in rapamycin-treated compared with untreated U1810 cells (Fig. 2b and Supplementary Fig. 8a). This approach unveiled a significant alteration of the U1810 transcriptome with the identification of 1,622 significantly (fold change > 1.5 or < 0.75 and $P < 0.001$) up- or downregulated genes (Fig. 2b, c and Supplementary Fig. 8a). A large fraction of the identified genes (141 genes; 8.7%) were related to autophagy (Fig. 2c and Supplementary Fig. 8b). There is an overall coincidence across the autophagy-related genes between the alteration of the GRO-seq signal and the absence of H4K16 acetylation. Indeed, 55 genes, that is 39% of the autophagy-related genes (including genes belonging to the autophagic core machinery) identified by GRO-seq analysis, showed reduced H4K16ac tag counts upon rapamycin treatment (Fig. 2c and Supplementary Fig. 9a, b).

To provide further evidence that the downregulation of H4K16ac during autophagy is part of a specific program, three extra histone modifications were examined, namely H3K4me3, H4K12ac and H4K8ac. In fact, whereas H4K16ac and H3K4me3 are known to be associated histone marks[21–23], K8 and K12 acetylation amounts on histone H4 are reported to be independent of H4K16 acetylation[13,24]. In agreement with the established molecular link between H4K16ac and H3K4me3, rapamycin-induced autophagy was associated with a reduction in H3K4me3 (Supplementary Fig. 10a, b), whereas H4K12ac and H4K18ac were left unaffected (Supplementary Fig. 10c). The joint downregulation of the H4K16ac and the H3K4me3 histone modifications was also observed in wild-type (WT) yeast, but not *atg1Δ*, *atg5Δ* and *atg7Δ* autophagy-deficient yeast or Sas2-overexpressing yeast (Supplementary 10d). Collectively, these genome-wide deep-sequencing analyses indicate that the observed deacetylation of H4K16 during autophagy results in transcriptional regulation of autophagy-related genes.

Because MTOR is involved in a wide variety of signalling pathways, treating cells with rapamycin or Torin 1 could cause the observed epigenetic changes by mechanisms unrelated to autophagy. To exclude this possibility, we tested the effect of rapamycin on the H4K16ac histone modification in *atg5*- and *atg7*-deficient MEF cells. These genes encode two ATG proteins that are essential for the canonical autophagy pathway (Fig. 1b, lower panels). Treatment of autophagy-deficient cells, namely $atg7^{-/-}$ or $atg5^{-/-}$ MEFs, with rapamycin did not lead to a similar degree of downregulation of H4K16ac (Fig. 1b, upper panels). Identical effects were observed in yeast (Supplementary Fig. 5).

Thus, the process of autophagy, independent of whether its induction required a SIRT1-dependent signalling pathway, was associated with deacetylation of H4K16. Collectively, these data suggest that alteration in another histone-modifying enzyme should be responsible for the observed modification in the acetylation status of H4K16. This observation prompted us to examine the status of hMOF during autophagy. Interestingly, although SIRT1 expression was not significantly altered upon rapamycin treatment (Supplementary Fig. 11a), hMOF expression was effectively downregulated in mammalian cells (Fig. 3a, b and Supplementary Figs 3a–c and 11a). Similarly, robust downregulation of hMOF expression was observed upon Torin 1 treatment or under amino-acid starvation (Supplementary Fig. 3a–d). Remarkably, the observed downregulation of hMOF upon rapamycin, Torin 1 or starvation treatment was abrogated when cells were co-treated with inhibitors of autophagy such as chloroquine (CQ), or 3-methyladenine (3MA) (Supplementary Fig. 3a–c). In yeast cells engineered to express an haemagglutinin (HA)-tagged version of the yeast homologue of hMOF, Sas2, rapamycin treatment induced a nearly complete loss of the HA signal within 3 h (Fig. 3g). Overexpression of Sas2 repressed the downregulation of H4K16ac upon rapamycin treatment in yeast cells (Fig. 3h). Altogether, these results showed that the downregulation of hMOF is part of the autophagy program.

The observed pronounced changes in amounts of H4K16 acetylation and associated transcriptional gene regulation suggested that there

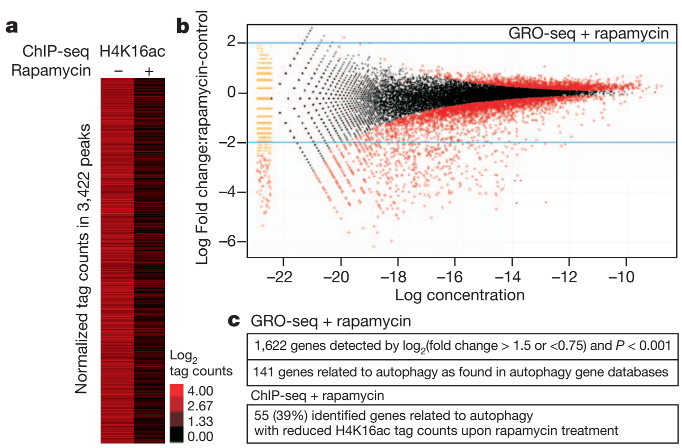


Figure 2 | Deacetylation of H4K16 by rapamycin treatment is associated with transcriptional regulation of autophagy-related genes. **a**, Heat map of H4K16ac ChIP-seq performed in U1810 cells without or with 8 h rapamycin treatment. Data are shown as $\log_2$ values of tag counts in the 3,422 regions defined as peaks in the control sample. **b**, We analysed *de novo* detection of transcripts using GRO-seq in 8 h rapamycin-treated U1810 cells and compared them with untreated U1810 cells. GRO-seq data visualized as 'MA' plots (log ratio versus abundance). The plot shows GRO-seq gene expression for pairwise comparison between rapamycin-treated versus control cells. The red points denote the differentially expressed genes. **c**, Autophagy-related genes identified as regulated by rapamycin in the GRO-seq data analysis and in the ChIP-seq data analysis.

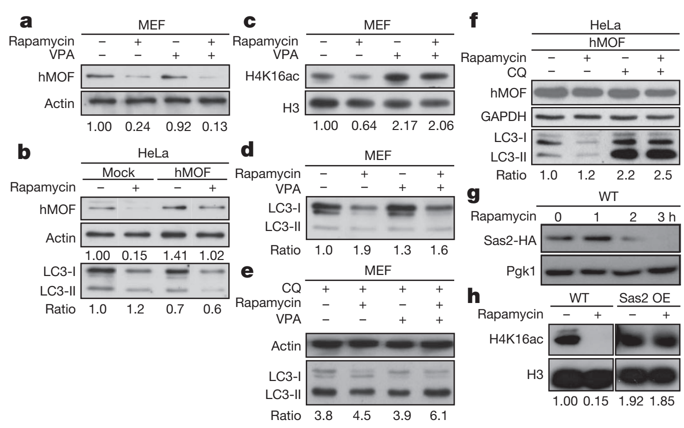


Figure 3 | Rapamycin-induced hMOF downregulation promotes deacetylation of H4K16. Rapamycin treatment (48 h) promoted the downregulation of the H4K16 histone acetyltransferase hMOF expression level in MEF (**a**) and transfected HeLa cells (**b**). VPA (1 mM) treatment counteracted rapamycin-induced H4K16ac downregulation (**c**) and decreased the LC3 ratio (**d**). **e**, Co-treatment with chloroquine (CQ, 10 μM) showed that the decrease in LC3 ratio was a result of an increase in autophagic flux. **f**, Inhibition of autophagy by CQ after hMOF overexpression shows that hMOF does not inhibit autophagic flux. **g**, The yeast homologue of hMOF, Sas2, tagged with 3× HA showed a complete disappearance of HA signal upon autophagy induction after 3 h. **h**, Overexpression of Sas2 repressed the downregulation of H4K16ac upon rapamycin treatment.

may be a functional role for this epigenetic change during autophagy. It has only recently become clear how histone modifications can play a regulatory role in apoptosis and how they can influence the decision between life and death (reviewed in ref. 25). A similar regulatory role for histone modifications could be present during autophagy. Shifting the equilibrium of hMOF and SIRT1 expression in favour of SIRT1 leads to a decrease in acetylation of H4K16 (ref. 13). Treatment with valproic acid (VPA) increased the acetylation status of H4K16 (Fig. 3c and Supplementary Fig. 11b–d) by reducing SIRT1 amounts[13]. Treatment with VPA was not only able to reverse rapamycin-induced downregulation of the H4K16ac histone modification (Fig. 3c), but also promoted LC3 conversion and turnover (Fig. 3d and Supplementary Fig. 11e). Administration of chloroquine to inhibit lysosomal activity enhanced the rapamycin-induced increase in LC3-II amounts. Co-treatment with VPA led to a further increase in the LC3-II amounts, confirming the increased autophagic flux in those cells (Fig. 3e and Supplementary Fig. 11f). Combined treatment with bafilomycin A_1 (BafA), which prevents maturation of autophagic vacuoles by inhibiting fusion between autophagosomes and lysosomes, led to similar results (Supplementary Fig. 12). The increase in the autophagic flux in those cells was further confirmed, making use of a tandem reporter construct, mRFP–GFP–LC3 (ref. 26) (Fig. 4a, b and Supplementary Figs 13a, 14 and 15). Rapamycin treatment resulted in an increase of yellow-colour-labelled LC3 puncta in U1810 and HeLa cells. In contrast, a remarkable increase in punctate red fluorescent signals was detected upon VPA and rapamycin treatment in both cell types (Fig. 4a, b and Supplementary Figs 13a and 14). Similarly, hMOF overexpression in HeLa and U1810 cells correlated with increased autophagic flux upon rapamycin treatment, as illustrated by LC3 immunoblotting (Fig. 3b, f) and the mRFP–GFP–LC3 autophagic flux assay (Supplementary Fig. 15).

We extended our investigation to the analysis of cell death. We observed a significant increase in cell death in human-derived cell lines co-treated with rapamycin, and VPA or the SIRT1-specific inhibitor Ex527 (ref. 13), as demonstrated by the appearance of condensed or fragmented nuclei (Fig. 4c, f and Supplementary Figs 13b, e and 16a, b). These results were further confirmed by fluorescence-activated cell sorting analysis of the appearance of a sub-G1 hypodiploid DNA peak (Supplementary Fig. 17a, b). In agreement with the observed effect of SIRT1 chemical inhibitors on the outcome of autophagy, we observed a significant increase in cell death in HeLa or U1810 cells upon SIRT1-knockdown and rapamycin treatment (Fig. 4e and Supplementary Fig. 13d). To investigate whether the observed cell death upon abrogation of H4K16 deacetylation after rapamycin treatment was a consequence of autophagy induction, we performed an extra set of experiments with CQ, BafA and 3MA, which inhibit different steps in the autophagic pathway. We noted that co-treatment with these inhibitors of autophagy was able to abrogate both rapamycin + VPA- and rapamycin + Ex527-induced cell death in human cancer cells (Fig. 4d, g, h and Supplementary Figs 13c and 16c–e). Knockdown of *ATG7* expression in mammalian cells prevented rapamycin + VPA-induced cell death and thus further strengthens the conclusion about an autophagy-mediated cell death (Supplementary Fig. 18). It is worth noting that the increased cell death upon VPA addition was not limited to rapamycin-induced autophagy, and was observed in amino-acid-starved HeLa cells (Fig. 4i, j). Furthermore, we investigated the link between hMOF activity and the outcome of autophagy. In agreement with the discovery that the perturbation of H4K16 acetylation status regulates the outcome of autophagy, we observed a significant increase in cell death in hMOF-overexpressing HeLa or U1810 cells upon rapamycin treatment, mimicking the effect of VPA/rapamycin co-treatment (Fig. 4f and Supplementary Fig. 13e). Collectively, these data indicate that the downregulation of hMOF, the associated reduction in H4K16 acetylation level and transcriptional regulation of autophagy-related genes are required for the proper progression of the autophagic process, and that disturbance of this epigenetic program results in cell death.

In conclusion, until now, nuclear events have not been considered of primary importance for autophagy, as enucleated cells are still able to accumulate GFP–LC3 puncta in response to autophagic stimuli[10]. Our data, however, unveil a critical link of the induction of autophagy

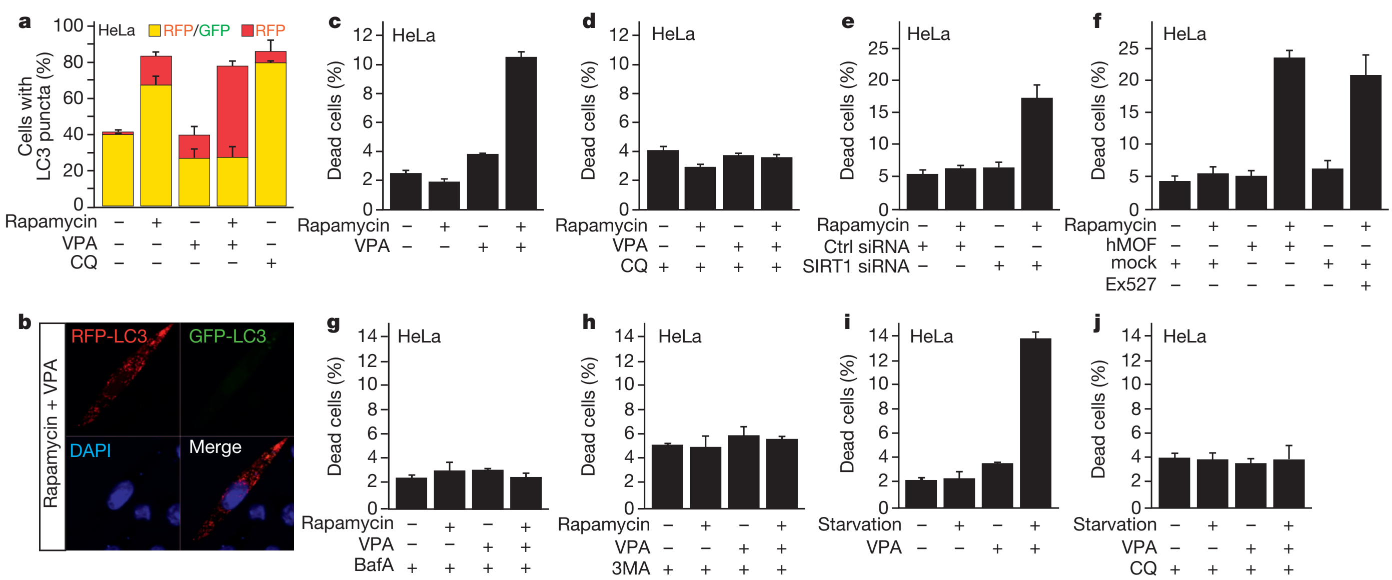


Figure 4 | Inhibition of H4K16ac downregulation upon autophagy induction results in cell death. **a**, VPA increases the autophagic flux in rapamycin-treated HeLa cells transfected with the mRFP–GFP–LC3 tandem reporter construct which allows distinction between autophagosomes (GFP+/RFP+ yellow puncta) and autolysosomes (GFP−/RFP+ red puncta). **b**, Confocal microscopy image of a cell treated with rapamycin and VPA depicting a high ratio of red to green LC3 puncta indicating an increase in autophagic flux. **c**, Co-treatment with VPA and rapamycin led to increased cell death. **d**, Co-treatment with CQ abrogated VPA + rapamycin-induced cell death. **e**, **f**, Increasing H4K16ac levels by either overexpression of hMOF, inhibition of SIRT1 by siRNA or the chemical inhibitor Ex527 promoted cell death upon rapamycin treatment. Co-treatment with the autophagy inhibitor bafilomycin A_1 (BafA, 40 nM) (**g**) or 3-methyladenine (3MA, 5 mM) (**h**) abrogated VPA + rapamycin-induced cell death. Treatment of HeLa cells upon amino-acid starvation with VPA induced cell death (**i**), which was rescued when cells were co-treated with CQ (**j**). Data are expressed as mean ± s.e.m. ($n = 3$).

and covalent histone H4K16 modifications with altered gene expression, including the regulation of key genes in the autophagy program. Our findings imply a molecular histone switch, where the balancing effects of hMOF and SIRT1 on H4K16 acetylation regulate autophagy (Supplementary Fig. 1). Our results do not oppose the findings about functionality of the autophagic process in enucleated cells, but add a new feedback regulatory network influencing the outcome of autophagy with respect to cell death/survival. The identification of tightly regulated histone modifications associated with the autophagic process offers an attractive conceptual framework to understand the short-term transcriptional response to stimuli eliciting autophagy, as well as constituting a potential aspect of long-term responses to autophagy.

METHODS SUMMARY

Antibodies and reagents used in this study are listed in Supplementary Tables 1 and 2. ON-TARGET plus SMARTpools short interfering RNAs (siRNAs) were purchased from Dharmacon (Supplementary Table 3). Experiments were performed on U1810, U2OS and HeLa human cancer cells and wild-type, *atg5*$^{-/-}$, *atg7*$^{-/-}$ and *sirt1*$^{-/-}$ mouse embryonic fibroblasts as well as wild-type, *atg1Δ*, *atg5Δ* and *atg7Δ* SEY6210 yeast cells. Histone protein extracts, total protein extracts and immunoblotting were performed as reported previously[13,27,28]. ChIP-seq and GRO-seq analyses were executed as described elsewhere[19,20]. Methods to monitor autophagy flux follow ref. 29. Statistical evaluations were performed by Student's *t*-test.

Full Methods and any associated references are available in the online version of the paper.

Received 7 September 2011; accepted 17 May 2013.

Published online 17 July; corrected online 21 August 2013 (see full-text HTML version for details).

1. Levine, B. & Kroemer, G. Autophagy in the pathogenesis of disease. *Cell* **132,** 27–42 (2008).
2. Mizushima, N., Levine, B., Cuervo, A. M. & Klionsky, D. J. Autophagy fights disease through cellular self-digestion. *Nature* **451,** 1069–1075 (2008).
3. Levine, B. & Yuan, J. Autophagy in cell death: an innocent convict? *J. Clin. Invest.* **115,** 2679–2688 (2005).
4. Yang, Z. & Klionsky, D. J. Eaten alive: a history of macroautophagy. *Nature Cell Biol.* **12,** 814–822 (2010).
5. Chen, Y. & Klionsky, D. J. The regulation of autophagy – unanswered questions. *J. Cell Sci.* **124,** 161–170 (2011).
6. Lee, I. H. & Finkel, T. Regulation of autophagy by the p300 acetyltransferase. *J. Biol. Chem.* **284,** 6322–6328 (2009).
7. Lee, I. H. *et al.* A role for the NAD-dependent deacetylase Sirt1 in the regulation of autophagy. *Proc. Natl Acad. Sci. USA* **105,** 3374–3379 (2008).
8. Yi, C. *et al.* Function and molecular mechanism of acetylation in autophagy regulation. *Science* **336,** 474–477 (2012).
9. Lin, S. Y. *et al.* GSK3–TIP60–ULK1 signaling pathway links growth factor deprivation to autophagy. *Science* **336,** 477–481 (2012).
10. Morselli, E. *et al.* Spermidine and resveratrol induce autophagy by distinct pathways converging on the acetylproteome. *J. Cell Biol.* **192,** 615–629 (2011).
11. Morselli, E. *et al.* Caloric restriction and resveratrol promote longevity through the Sirtuin-1-dependent induction of autophagy. *Cell Death Dis.* **1,** e10 (2010).
12. Vaquero, A., Sternglanz, R. & Reinberg, D. NAD+-dependent deacetylation of H4 lysine 16 by class III HDACs. *Oncogene* **26,** 5505–5520 (2007).
13. Hajji, N. *et al.* Opposing effects of hMOF and SIRT1 on H4K16 acetylation and the sensitivity to the topoisomerase II inhibitor etoposide. *Oncogene* **29,** 2192–2204 (2010).
14. Taipale, M. *et al.* hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol. Cell. Biol.* **25,** 6798–6810 (2005).
15. Smith, E. R. *et al.* A human protein complex homologous to the *Drosophila* MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Mol. Cell. Biol.* **25,** 9175–9188 (2005).
16. Kouzarides, T. Chromatin modifications and their function. *Cell* **128,** 693–705 (2007).
17. Shogren-Knaak, M. *et al.* Histone H4–K16 acetylation controls chromatin structure and protein interactions. *Science* **311,** 844–847 (2006).
18. Kind, J. *et al.* Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in *Drosophila*. *Cell* **133,** 813–828 (2008).
19. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322,** 1845–1848 (2008).
20. Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474,** 390–394 (2011).
21. Ruthenburg, A. J. *et al.* Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell* **145,** 692–706 (2011).
22. Wang, Z. *et al.* Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138,** 1019–1031 (2009).
23. Katoh, H. *et al.* FOXP3 orchestrates H4K16 acetylation and H3K4 trimethylation for activation of multiple genes by recruiting MOF and causing displacement of PLU-1. *Mol. Cell* **44,** 770–784 (2011).
24. Zhou, Y. & Grummt, I. The PHD finger/bromodomain of NoRC interacts with acetylated histone H4K16 and is sufficient for rDNA silencing. *Curr. Biol.* **15,** 1434–1438 (2005).
25. Fullgrabe, J., Hajji, N. & Joseph, B. Cracking the death code: apoptosis-related histone modifications. *Cell Death Differ.* **17,** 1238–1243 (2010).
26. Kimura, S., Noda, T. & Yoshimori, T. Dissection of the autophagosome maturation process by a novel reporter protein, tandem fluorescent-tagged LC3. *Autophagy* **3,** 452–460 (2007).
27. Shechter, D., Dormann, H. L., Allis, C. D. & Hake, S. B. Extraction, purification and analysis of histones. *Nature Protocols* **2,** 1445–1457 (2007).
28. Burguillos, M. A. *et al.* Caspase signalling controls microglia activation and neurotoxicity. *Nature* **472,** 319–324 (2011).
29. Klionsky, D. J. *et al.* Guidelines for the use and interpretation of assays for monitoring autophagy. *Autophagy* **8,** 445–544 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. Mc Inerney, M. Malewicz, S. Orrenius and T. Panaretakis for discussion, and L. Guarente, V. Kaminskyy, M. Komatsu, G. Mc Inerney, M. Panas, R. G. Roeder and L. Xiaoling for reagents and cell lines. J.F. is supported by a fellowship from the Karolinska Institutet Foundations, M.A.L.-D. is partly supported by a Rackham Predoctoral Fellowship and W.L. is supported by breast cancer research Postdoctoral Fellowship Award (BC110381) from the US Department of Defense. This work was supported by a National Institutes of Health grant GM53396 (to D.J.K.) and National Institutes of Health/National Cancer Institute and Department of Defense grants (to M.G.R.), the Swedish Cancer Society, the Swedish Childhood Cancer Foundation (to B.J. and O.H.) and the Swedish Research Council (to B.J.). M.G.R. is an investigator of the Howard Hughes Medical Institute.

Author Contributions J.F., R.B.S. and B.J. performed experiments in mammalian cells. M.A.L.-D. performed yeast experiments. N.H. performed ChIP-seq. W.L. performed GRO-seq. N.H. and Q.M. performed bioinformatical analysis. J.F., O.H., M.G.R., D.J.K. and B.J. designed the study, and analysed and interpreted the data. The first draft of the paper was written by J.F. and B.J. All authors discussed the results and commented on or edited the manuscript. D.J.K. and B.J. share senior authorship of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.J. (bertrand.joseph@ki.se).

METHODS

Cell culture and transfection. Non-small-cell lung carcinoma U1810 cells, osteosarcoma U2OS cells, cervical cancer HeLa cells and wild-type, *atg5*$^{-/-}$, *atg7*$^{-/-}$ and *sirt1*$^{-/-}$ mouse embryonic fibroblasts were cultured using standard procedures[13,30]. The *atg5*$^{-/-}$ and WT MEF cell lines were gifts from G. Mc Inerney, the *atg7*$^{-/-}$ MEF cell line was a gift from M. Komatsu[30] and the *sirt1*$^{-/-}$ MEF cell line was a gift from X. Li. SEY6210 wild-type, *atg1Δ*, *atg5Δ* and *atg7Δ* yeast cells were grown as described previously[31]. Reagents used in this study are listed in Supplementary Table 2. Plasmids encoding HA-human MOF, Flag–human SIRT1 and mRFP–GFP–LC3 were gifts from R. G. Roeder, L. Guarente and G. Mc Inerney, respectively. Transfection of cells used lipofectamine and lipofectamine reagent in HeLa cells and X-tremeGENE HP DNA transfection reagent in U1810 cells.

Histone extracts and immunoblotting. Histone protein extracts were performed as described elsewhere[13,27] using TCA precipitation and H_2SO_4 extraction or using the Histone Purification Mini Kit (Active Motif). Total protein extracts and immunoblotting were performed as reported previously[28]. GAPDH, β-actin, histone 3 (H3) and phosphoglycerate kinase 1 (Pgk1) were used as standards for equal loading of protein. Antibodies used in this study are listed in Supplementary Table 1. Densitometry was done using ImageJ.

Yeast procedures. The GFP–Atg8 processing assay and fluorescence microscopy were performed as described previously[32]. If not stated otherwise, all experiments were performed at 1 h treatment.

Immunofluorescence and confocal microscopy. For confocal microscopy analysis, the adherent mammalian cells were grown on coverslips. Paraformaldehyde-fixed cells were blocked in HEPES, 3% bovine serum albumin, 0.3% Triton X-100 and incubated with primary (4 °C, overnight) and secondary (room temperature, 1 h) antibodies. Samples were mounted with Vectashield (Vector Laboratories) and analysed with Zeiss 510 META confocal laser scanning microscopy (Zeiss)[28].

Evaluation of autophagic flux. Autophagy was assessed following the 'Guidelines for the use and interpretation of assays for monitoring autophagy'[29]. To monitor the autophagic flux, a tandem reporter construct mRFP–GFP–LC3 was used[26]. The green fluorescence of this tandem reporter is attenuated in the acidic pH lysosomal environment, whereas the mRFP is not. Therefore, the green fluorescent component of the composite yellow fluorescence from this mRFP–GFP–LC3 reporter is lost upon autophagosome fusion with a lysosome, whereas the red fluorescence remains detectable. Thus this probe allows distinction between autophagosomes (GFP+/RFP+ yellow puncta) and autolysosomes (GFP−/RFP+ red puncta). At 24 h after plating, the cells were transfected with the mRFP–GFP–LC3 plasmid, alone or in combination with *SIRT1* siRNA, control siRNA or *hMOF* plasmid. ON-TARGET plus SMARTpools siRNAs were purchased from Dharmacon (Supplementary Table 3). The next day, cells were treated for an extra 24 h with the indicated compound(s). Cells were then fixed using 4% paraformaldehyde, and autophagy was determined by quantification of the number of cells with LC3-positive organelles, counting at least 100 cells in triplicate per condition. The presence of autophagic vacuoles expressing endogenous LC3 was also assessed.

Cell death quantification. After treatment, cells were fixed in 4% paraformaldehyde, collected and cytospins were prepared. Subsequently, DNA was stained with Hoechst 33342 (0.1 mg ml^{-1}; Molecular Probes/Invitrogen). The number of dying cells was measured quantitatively by assessing the percentage of cells with fragmented, damaged or condensed nuclei.

Fluorescence-activated cell sorting analysis. Quantification of PI (Sigma) and TMRE (Molecular Probes/Invitrogen) staining was performed with a FACSCalibur flow cytometer (Becton Dickinson) using standard procedures[28].

ChIP-sequencing. For ChIP-seq analysis, 5 μg of chromatin was used in two separate immunoprecipitations and combined in one elution for each condition. Subsequently, the DNA sequencing library was made using a kit from Illumina (catalogue number 1003473) except that Illumina TruSeq adaptors (to enable multiplexing) were used. The library was analysed by Solexa/Illumina Hi-seq. After pre-filtering the raw data by removing sequenced adapters and low quality reads, the sequence tags were aligned to the human genome (assembly hg19) with the Bowtie alignment tool[33]. To avoid any PCR-generated spikes we allowed only one read per chromosomal position, thus eliminating PCR bias. From the filtered raw data, 8 million unique reads per sample were used for peak detection. Peak detection was performed using the CisGenome program[34] with a two-sample analysis where sequenced input (1%) was used as a negative control. Peaks were called with a window statistic cut-off of 3 and a $\log_2$-fold change of 2. Using the defined chromosomal peak regions from the no-treatment condition, the number of tags were counted in the corresponding rapamycin-treated sample and heat maps were generated using Java Treeview[35].

GRO-sequencing. GRO-seq experiments were performed as previously reported[19,20]. Briefly, cells were washed with cold 1× PBS buffer and swelled in swelling buffer (10 mM Tris-Cl pH 7.5, 2 mM $MgCl_2$, 3 mM $CaCl_2$) for 5 min on ice and collected. Cells were lysed in lysis buffer (swelling buffer with 0.5% NP-40, 2 units per millilitre Superase In and 10% glycerol) and finally re-suspended in 100 μl of freezing buffer (50 mM Tris-Cl pH 8.3, 40% glycerol, 5 mM $MgCl_2$, 0.1 mM EDTA). For the run-on assay, re-suspended nuclei were mixed with an equal volume of reaction buffer (10 mM Tris-Cl pH 8.0, 5 mM $MgCl_2$, 1 mM DTT, 300 mM KCl, 20 units of SUPERase In, 1% sarkosyl, 500 μM ATP, GTP, and Br-UTP, 2 μM CTP) and incubated for 5 min at 30 °C. The nuclear-run-on RNA (NRO-RNA) was then extracted with TRIzol LS reagent (Invitrogen) following the manufacturer's instructions. NRO-RNA was then subjected to base hydrolysis on ice for 40 min and followed by treatment with DNase I and antarctic phosphatase. To purify the Br-UTP-labelled nascent RNA, the NRO-RNA was immunoprecipitated with anti-BrdU argarose beads (Santa Cruz Biotech) in binding buffer (0.5× SSPE, 1 mM EDTA, 0.05% Tween-20). To repair the end, the immunoprecipitated BrU-RNA was re-suspended in a 50 μl reaction (45 μl DEPC water, 5.2 μl T4 PNK buffer, 1 μl SUPERase In and 1 μl T4 PNK (NEB)) and incubated at 37 °C for 1 h. The RNA was extracted and precipitated using acidic phenol-chloroform. The complementary DNA (cDNA) synthesis was performed basically as in ref. 36 with a few minor modifications. First, RNA fragments were subjected to a poly-A tailing reaction by poly-A polymerase (NEB) for 30 min at 37 °C. Subsequently, reverse transcription was performed using oNTI223 primer (5′-pGATCGTCGGACTGTAGAACTCT; CAAGCAGAAGACGGCATACGA TTTTTTTTTTTTTTTTTTTTVN). Second, tailed RNA (8.0 μl) was subjected to reverse transcription using superscript III (Invitrogen). The cDNA products were separated on a 10% polyacrylamide TBE-urea gel and the extended first-strand product (100–500 base pairs) was excised and recovered by gel extraction. After that, the first-strand cDNA was circularized by CircLigase (Epicentre) and relinearized by Ape1 (NEB). Relinearized single strand cDNA (sscDNA) was separated on a 10% polyacrylamide TBE gel and the product of needed size was excised (~120–320 base pairs) for gel extraction. Finally, single strand cDNA template was amplified by PCR using the Phusion High-Fidelity enzyme (NEB) according to the manufacturer's instructions with two oligonucleotide primers, oNTI200 (5′-CAAGCAGAAGACGGCATA) and oNTI201 (5′-AATGATACGG CGACCACCGACAGGT TCAGAGTTCTACAGTCCGACG). DNA was then sequenced on the Illumina HiSeq2000 according to the manufacturer's instructions, using the small RNA sequencing primer 5′-CGACAGGTTCAGAGTTCTA CAGTCCGACGATC.

Database analyses. The following autophagy databases were used to indentify autophagy-related genes: the human autophagy database at http://autophagy.lu/ and the autophagy database at http://www.tanpaku.org/autophagy/overview.html.

Statistical analyses. Bars and error bars represent mean with s.e.m. Statistical evaluations were performed by Student's *t*-test.

30. Komatsu, M. *et al.* Impairment of starvation-induced and constitutive autophagy in Atg7-deficient mice. *J. Cell Biol.* **169,** 425–434 (2005).
31. Cao, Y., Cheong, H., Song, H. & Klionsky, D. J. In vivo reconstitution of autophagy in *Saccharomyces cerevisiae*. *J. Cell Biol.* **182,** 703–713 (2008).
32. Shintani, T. & Klionsky, D. J. Cargo proteins facilitate the formation of transport vesicles in the cytoplasm to vacuole targeting pathway. *J. Biol. Chem.* **279,** 29889–29894 (2004).
33. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
34. Ji, H. *et al.* An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nature Biotechnol.* **26,** 1293–1300 (2008).
35. Saldanha, A. J. Java Treeview–extensible visualization of microarray data. *Bioinformatics* **20,** 3246–3248 (2004).
36. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weisman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324,** 218–223 (2009).

doi:10.1038/nature12466

Optical control of mammalian endogenous transcription and epigenetic states

Silvana Konermann[1,2]*, Mark D. Brigham[1,2,3]*, Alexandro E. Trevino[1,2], Patrick D. Hsu[1,2,4], Matthias Heidenreich[1,2], Le Cong[1,2,5], Randall J. Platt[1,2], David A. Scott[1,2], George M. Church[1,6] & Feng Zhang[1,2]

The dynamic nature of gene expression enables cellular programming, homeostasis and environmental adaptation in living systems. Dissection of causal gene functions in cellular and organismal processes therefore necessitates approaches that enable spatially and temporally precise modulation of gene expression. Recently, a variety of microbial and plant-derived light-sensitive proteins have been engineered as optogenetic actuators, enabling high-precision spatiotemporal control of many cellular functions[1–11]. However, versatile and robust technologies that enable optical modulation of transcription in the mammalian endogenous genome remain elusive. Here we describe the development of light-inducible transcriptional effectors (LITEs), an optogenetic two-hybrid system integrating the customizable TALE DNA-binding domain[12–14] with the light-sensitive cryptochrome 2 protein and its interacting partner CIB1 from *Arabidopsis thaliana*. LITEs do not require additional exogenous chemical cofactors, are easily customized to target many endogenous genomic loci, and can be activated within minutes with reversibility[6,15]. LITEs can be packaged into viral vectors and genetically targeted to probe specific cell populations. We have applied this system in primary mouse neurons, as well as in the brain of freely behaving mice *in vivo* to mediate reversible modulation of mammalian endogenous gene expression as well as targeted epigenetic chromatin modifications. The LITE system establishes a novel mode of optogenetic control of endogenous cellular processes and enables direct testing of the causal roles of genetic and epigenetic regulation in normal biological processes and disease states.

The LITE system uses a modular design consisting of two independent components (Fig. 1a). The first component is the genomic anchor and includes a customizable DNA-binding domain, based on transcription activator-like effectors (TALEs)[12,13] from *Xanthomonas* sp., fused to the light-sensitive cryptochrome 2 (CRY2) protein from *Arabidopsis thaliana*[6,15] (TALE–CRY2). The second component includes the interacting partner of CRY2, CIB1[6,15], fused to a desired effector domain (CIB1–effector). In the absence of light (inactive state), TALE–CRY2 binds the promoter region of the target gene while CIB1–effector remains free within the nuclear compartment. Illumination with blue light triggers a conformational change in CRY2 and subsequently recruits CIB1–effector (VP64 shown in Fig. 1a) to the target locus to mediate transcriptional modulation. This modular design allows each LITE component to be independently engineered, allowing the same genomic anchor to be combined with activating or repressing effectors[16,17] to exert positive and negative transcriptional control over the same endogenous genomic locus. In principle, the genomic anchor may also be replaced with other DNA-binding domains such as zinc-finger proteins[16] or RNA-guided DNA-binding domains based on nucleolytically inactive mutants of Cas9 (Extended Data Fig. 1)[18–22].

To identify the most effective architecture, we assessed the efficacy of different LITE designs by measuring transcriptional changes of the neural lineage-specifying transcription factor neurogenin 2 (*Neurog2*) induced by blue light illumination (Fig. 1b). Three out of four initial LITE pairings produced significant light-induced *Neurog2* messenger RNA upregulation in Neuro 2a cells ($P < 0.001$, Fig. 1b). Of these, TALE(*Neurog2*)–CRY2PHR complexed with CIB1–VP64 (TALE(*Neurog2*)–CRY2PHR/

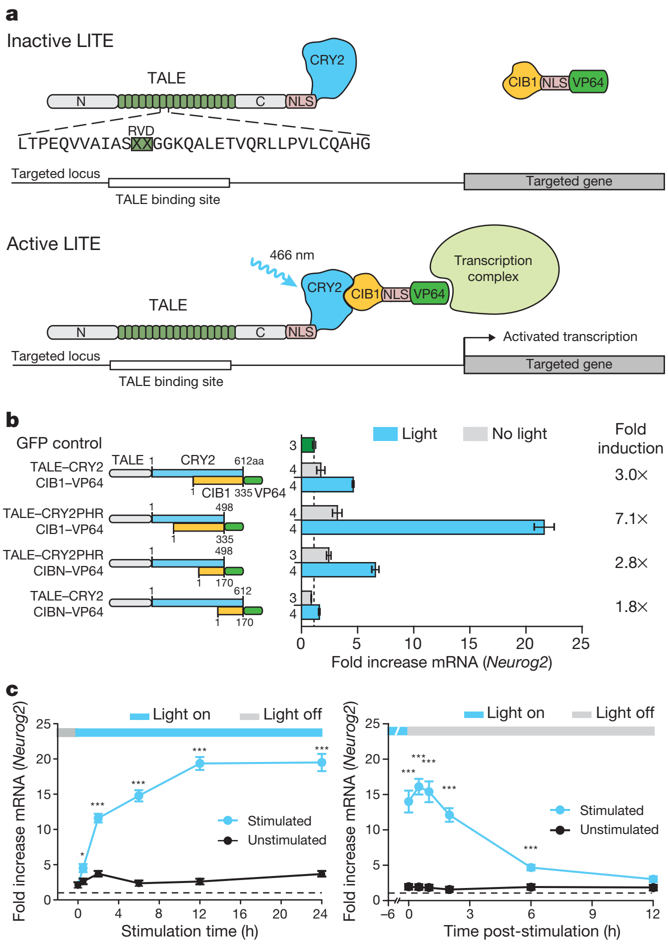


Figure 1 | Design and optimization of the LITE system. **a**, Schematic of the LITE system. Light stimulation induces dimerization of CRY2 and CIB1, recruiting the effector to the target promoter. **b**, LITE architecture was optimized by fusing TALE and the transcriptional activator VP64[14,16] to different truncations of CRY2 and CIB1[6] (*n* next to each bar). **c**, Time-course of light-dependent *Neurog2* upregulation and decay post-illumination ($n = 4$ biological replicates; $*P < 0.05$; $***P < 0.001$). Cells were stimulated with 5 mW cm^{-2} light (466 nm, 1 s pulses at 0.067 Hz). Mean ± s.e.m. in all panels.

[1]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [2]McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA. [4]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [5]Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, Massachusetts 02115, USA. [6]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.
*These authors contributed equally to this work.

CIB1–VP64), containing a truncated CRY2 consisting of the photolyase homology region alone (CRY2PHR[6], amino acids 1–498), yielded strongest light-mediated transcription activation as well as the highest induction ratio (light/no light mRNA levels). Therefore TALE–CRY2PHR/CIB1–VP64 was used in subsequent experiments. To ensure optimal function, we also systematically tuned light stimulation parameters (wavelength[23], Extended Data Fig. 2a; duty cycle, Extended Data Fig. 2b; light intensity, Extended Data Fig. 2c, d and Supplementary Discussion; and choice of activation domain, Extended Data Fig. 2e).

Although the interaction between CRY2 and CIB1 occurs on a subsecond timescale[6], LITE-mediated transcriptional activation is likely to be dependent on many factors, including rate of transcription, mRNA processing and transcript stability[24,25]. We found that LITE-mediated *Neurog2* expression increased considerably as early as 30 min after initial stimulation and rose steadily until saturating at 12 h with approximately 20-fold upregulation compared to green fluorescent protein (GFP)-transfected negative controls (Fig. 1c). Interestingly, *Neurog2* transcript levels continued to increase for up to 30 min post-illumination, an effect that may have resulted from residual CRY2PHR–CIB1 dimerization, which has been shown to reach complete dissociation after 15 min (ref. 6), or from previously recruited RNA polymerases. Thereafter, *Neurog2* mRNA returned to baseline levels with a half-life of ~3 h. In contrast, a small-molecule inducible TALE system based on the plant hormone abscisic acid receptor[26] showed slower on- and off-kinetics (Extended Data Fig. 3).

To apply LITEs to neuronal applications, we developed an adeno-associated virus (AAV)-based vector (Fig. 2a, b) for the delivery of TALE genes and a simplified process for AAV production (Extended Data Fig. 4 and Methods). The single-stranded DNA-based genome of AAV is less susceptible to recombination, providing an advantage over lentiviral vectors[27]. We evaluated a panel of 28 TALE activators targeting the mouse genome in primary neurons and found that most were able to upregulate transcription in primary neurons (Fig. 2c). Moreover, *in vivo* expression of TALE(*Grm2*)–VP64 in the prefrontal cortex (PFC) (Fig. 2d, e) induced a 2.5-fold increase in *Grm2* mRNA levels compared to GFP-only controls (Fig. 2f).

Similarly, we introduced LITEs into primary cortical neurons via co-delivery of two AAVs (Fig. 3a, b). We tested a *Grm2*-targeted LITE at two light pulsing frequencies with a reduced duty cycle of 0.8% to ensure neuron health (Extended Data Fig. 5a). Both stimulation conditions achieved an approximately sevenfold light-dependent increase in *Grm2* mRNA levels (Fig. 3c). Further study verified that substantial target gene expression increases could be attained quickly (fourfold upregulation of mRNA within 4 h; Fig. 3d). In addition, we observed significant upregulation of mGluR2 protein after stimulation, confirming that LITE-mediated transcriptional changes are translated to the protein level ($P < 0.01$ vs GFP control, $P < 0.05$ vs no-light condition; Fig. 3e). To test the *in vivo* functionality of the LITE system, we stereotactically delivered a 1:1 mixture of high-titre AAV vectors carrying TALE(*Grm2*)–CIB1 and CRY2PHR–VP64 into the PFC. We used a previously established fibre optic cannula system to deliver light to LITE-expressing neurons *in vivo* (Fig. 3f, g and Extended Data Fig. 5b)[28]. After 12 h of stimulation, we observed a significant increase in *Grm2* mRNA compared with unstimulated PFC (Fig. 3h, $P \leq 0.01$). Taken together, these results confirm that LITEs enable optical control of endogenous gene expression in cultured neurons and *in vivo*.

Given persistent baseline upregulation *in vivo*, we undertook further rounds of optimization to reduce background activity and improve the gene induction ratio. We observed that TALE(*Grm2*)–CIB1 alone produced similar levels of upregulation as LITE background activation, whereas CRY2PHR–VP64 alone did not significantly affect transcription (Extended Data Fig. 5c). Therefore we rationalized that LITE-dependent background transcriptional activation arises mainly from TALE–CIB1.

We subsequently designed a comprehensive screen to reduce baseline TALE–CIB1-mediated upregulation, focusing on two strategies. First, although CIB1 is a plant transcription factor, it may have

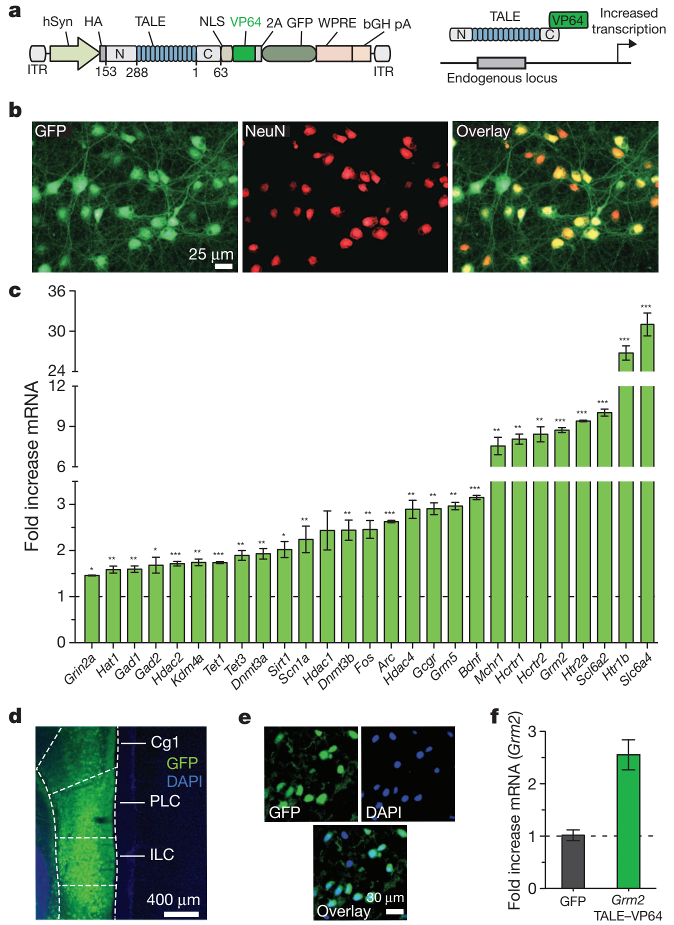


Figure 2 | *In vitro* and *in vivo* AAV-mediated TALE delivery targeting endogenous loci in neurons. **a**, Schematic AAV vectors for TALE delivery. **b**, Representative images of primary cortical neurons expressing TALE–VP64. **c**, TALE–VP64 constructs targeting a variety of endogenous neuronal genes were screened for transcriptional activation in primary cortical neurons (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; $n = 3$ biological replicates). **d**, TALE–VP64 expression in PFC. DAPI, 4′,6-diamidino-2-phenylindole; Cg1, cingulate cortex area 1; PLC, prelimbic cortex; ILC, infralimbic cortex. **e**, Higher magnification image of TALE–VP64-expressing neurons in PFC. **f**, *Grm2* mRNA upregulation by TALE–VP64 *in vivo* in PFC ($n = 4$ animals). Mean ± s.e.m. in all panels.

intrinsic activity in mammalian cells[15]. To address this, we deleted three CIB1 regions conserved amongst basic helix–loop–helix transcription factors of higher plants (Extended Data Fig. 6). Second, to prevent TALE–CIB1 from binding the target locus in absence of light, we engineered TALE–CIB1 to localize in the cytoplasm pending light-induced dimerization with the nuclear localization signal (NLS)-containing CRY2PHR–VP64 (Extended Data Fig. 7a and b). To test both strategies independently or in combination, we evaluated 73 distinct LITE architectures and identified 12 effector-targeting domain pairs (denoted by the '+' column in Extended Data Fig. 6) with both improved light-induction efficiency and reduced background (fold mRNA increase in the no-light condition compared with the original LITE; $P < 0.05$). One architecture successfully incorporating both strategies, designated LITE2.0, demonstrated the strongest light induction (light/no-light = 20.4) and resulted in greater than sixfold reduction of background activation compared with the original design (Fig. 3i). Another, LITE1.9.1, produced minimal background activation (1.06) while maintaining fourfold light induction (Extended Data Fig. 7c).

Finally, we sought to expand the range of transcriptional processes addressable by TALEs and LITEs. We reasoned that TALE-mediated targeting of histone effectors to endogenous loci could induce specific

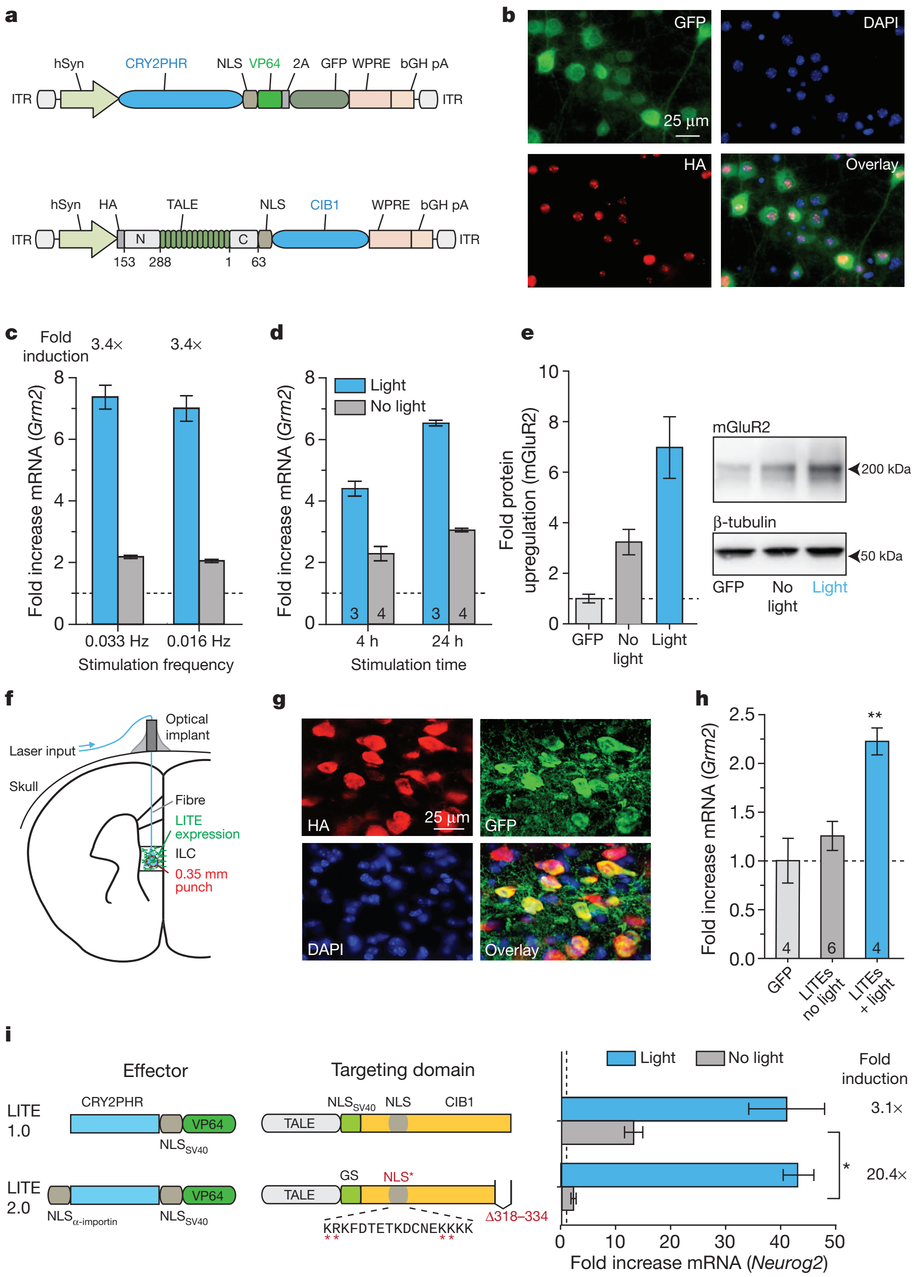


Figure 3 | LITE-mediated optogenetic modulation of endogenous transcription in primary neurons and *in vivo*. **a**, Schematic of AAV-LITE constructs. **b**, Images of primary neurons expressing LITE constructs. HA, haemagglutinin tag. **c**, Light-induced activation of *Grm2* in primary neurons after 24 h of stimulation (250 ms pulses at 0.033 Hz or 500 ms pulses at 0.016 Hz; 5 mW cm^{-2}; $n = 4$ biological replicates). **d**, Upregulation of *Grm2* in primary cortical neurons after 4 h or 24 h of stimulation. Expression levels are shown relative to neurons transduced with GFP only (number of biological replicates denoted within graph bars). **e**, Light-mediated changes in mGluR2 protein levels ($n = 7$ biological replicates). **f**, Schematic of *in vivo* optogenetic stimulation setup. **g**, Representative images of PFC neurons expressing both LITE components. **h**, Light-induced activation of endogenous *Grm2* expression using LITEs transduced into ILC. (**$P < 0.01$; number of animals denoted within graph bars.) **i**, LITE2.0 significantly reduces the level of background activation in Neuro 2a cells ($n = 3$ biological replicates). Mean ± s.e.m. in all panels.

epigenetic modifications, which would enable the interrogation of epigenetic as well as transcriptional dynamics (Fig. 4a)[29]. We fused CRY2PHR with four concatenated mSin3 interaction domains (SID4X; Fig. 4b and Extended Data Fig. 8) and observed light-mediated transcriptional repression of *Grm2* in neurons (Fig. 4c) accompanied by an approximately twofold reduction in H3K9 acetylation at the targeted *Grm2* promoter (Fig. 4d). To expand the diversity of histone residue targets for locus-specific histone modification, we next derived a set of 32 repressive histone effector domains (Supplementary Tables 1–5). Selected from across a wide phylogenetic spectrum, the domains include histone deacetylases (HDACs), methyltransferases (HMTs), acetyltransferase (HAT) inhibitors, as well as HDAC and HMT recruiting proteins. Preference was given to proteins and functional truncations of small size to facilitate efficient AAV packaging. The resulting epigenetic mark-modifying TALE–histone effector fusion constructs (epiTALEs) were evaluated in primary neurons and Neuro 2a cells for their ability to repress *Grm2* and *Neurog2* transcription, respectively (Fig. 4e, f and Extended Data Fig. 9). In primary neurons, 23 out of 24 epiTALEs successfully repressed transcription of *Grm2* ($P < 0.05$). Similarly, epiTALE expression in Neuro 2a cells led to decreased *Neurog2* expression for 20 of the 32 histone effector domains tested (Extended Data Fig. 9; $P < 0.05$). We then expressed a subset of promising epiTALEs in primary neurons and Neuro 2a cells and quantified the relative histone residue mark levels at the target locus using chromatin immunoprecipitation

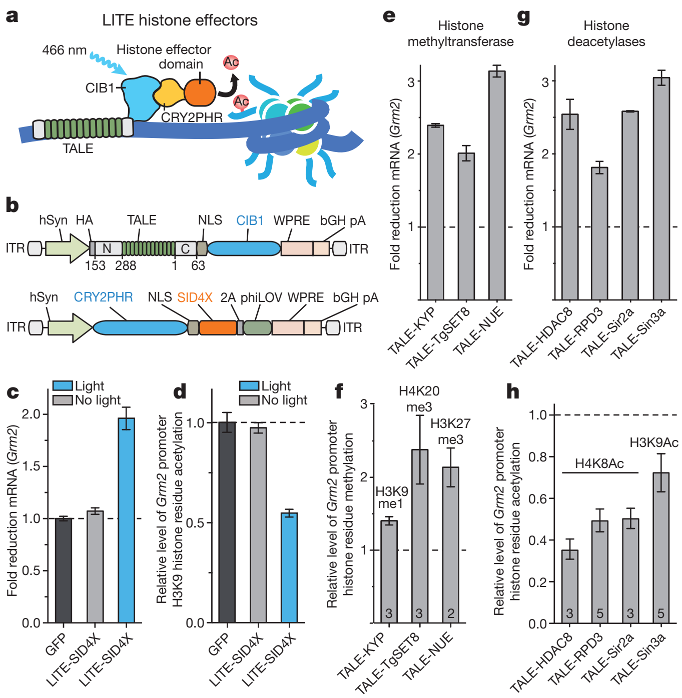


Figure 4 | TALE- and LITE-mediated epigenetic modifications. **a**, LITE epigenetic modifiers (epiLITE). **b**, epiLITE AAV vectors. **c**, epiLITE-mediated repression of endogenous *Grm2* in neurons ($n = 4$ biological replicates). **d**, epiLITE-mediated decrease in H3K9 histone acetylation at the *Grm2* promoter ($n = 4$ biological replicates). **e**, **f**, epiTALE methyltransferases mediated decrease in *Grm2* mRNA and corresponding enrichment of H3K9me1, H4K20me3 and H3K27me3 at the *Grm2* promoter ($n = 3$ biological replicates). **g**, **h**, epiTALE histone deacetylases mediated repression of *Grm2* and corresponding decreases in H4K8Ac and H3K9Ac marks at the *Grm2* promoter (n denoted within graph). Mean ± s.e.m. in all panels.

followed by reverse transcription and quantitative PCR (ChIP-qRT–PCR; Fig. 4g, h and Extended Data Fig. 10). In primary neurons and Neuro 2a cells, epiTALE-mediated modifications were observed for the following histone marks: H3K9me1 (KYP (*A. thaliana*)), H4K20me3 (TgSET8 (*Toxoplasma gondii*)), H3K27me3 (NUE and PHF19 (*Chlamydia trachomatis* and *Homo sapiens*)), H3K9ac (Sin3a, Sirt3 and NcoR (all *H. sapiens*)) and H4K8ac (HDAC8, RPD3 and Sir2a (*Xenopus laevis*, *Saccharomyces cerevisiae* and *Plasmodium falciparum*)). These domains provide a ready source of effectors for LITE-mediated control of specific epigenetic modifications.

Spatiotemporally precise perturbation of transcription and epigenetic states *in vivo* using LITE can enable researchers to test the causal role of gene regulation in diverse processes including development, learning and disease. TALEs can be conveniently customized to target a wide range of genomic loci, and other DNA-binding domains such as the RNA-guided Cas9 enzymes may be used in lieu of TALE to enable multiplexed transcriptional and epigenetic engineering of individual or groups of genomic loci in cells and whole organisms[18–20]. Novel modes of LITE modulation can also be achieved by replacing the effector module with functional domains such as chromatin-modifying enzymes[29]. The LITE system enables a powerful set of novel capabilities for the optogenetic toolbox and establishes a highly generalizable and versatile platform for reverse-engineering the function and regulation of mammalian genomes.

METHODS SUMMARY

LITE constructs were transfected into Neuro 2a cells using GenJet. AAV vectors carrying TALE or LITE constructs were used to transduce mouse primary embryonic cortical neurons as well as the mouse brain *in vivo*. RNA was extracted and reverse transcribed and mRNA levels were measured using TaqMan-based qRT–PCR. Light emitting diodes or solid-state lasers were used for light delivery in tissue culture and *in vivo*, respectively.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 February; accepted 16 July 2013.

Published online 23 July 2013.

1. Deisseroth, K. Optogenetics. *Nature Methods* **8,** 26–29 (2011).
2. Zhang, F. *et al.* The microbial opsin family of optogenetic tools. *Cell* **147,** 1446–1457 (2011).
3. Levskaya, A., Weiner, O. D., Lim, W. A. & Voigt, C. A. Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature* **461,** 997–1001 (2009).
4. Yazawa, M., Sadaghiani, A. M., Hsueh, B. & Dolmetsch, R. E. Induction of protein-protein interactions in live cells using light. *Nature Biotechnol.* **27,** 941–945 (2009).
5. Strickland, D. *et al.* TULIPs: tunable, light-controlled interacting protein tags for cell biology. *Nature Methods* **9,** 379–384 (2012).
6. Kennedy, M. J. *et al.* Rapid blue-light-mediated induction of protein interactions in living cells. *Nature Methods* **7,** 973–975 (2010).
7. Shimizu-Sato, S., Huq, E., Tepperman, J. M. & Quail, P. H. A light-switchable gene promoter system. *Nature Biotechnol.* **20,** 1041–1044 (2002).
8. Ye, H., Daoud-El Baba, M., Peng, R. W. & Fussenegger, M. A synthetic optogenetic transcription device enhances blood-glucose homeostasis in mice. *Science* **332,** 1565–1568 (2011).
9. Polstein, L. R. & Gersbach, C. A. Light-inducible spatiotemporal control of gene activation by customizable zinc finger transcription factors. *J. Am. Chem. Soc.* **134,** 16480–16483 (2012).
10. Bugaj, L. J., Choksi, A. T., Mesuda, C. K., Kane, R. S. & Schaffer, D. V. Optogenetic protein clustering and signaling activation in mammalian cells. *Nature Methods* **10,** 249–252 (2013).
11. Zhang, F. *et al.* Multimodal fast optical interrogation of neural circuitry. *Nature* **446,** 633–639 (2007).
12. Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326,** 1509–1512 (2009).
13. Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326,** 1501 (2009).
14. Zhang, F. *et al.* Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nature Biotechnol.* **29,** 149–153 (2011).
15. Liu, H. *et al.* Photoexcited CRY2 interacts with CIB1 to regulate transcription and floral initiation in *Arabidopsis*. *Science* **322,** 1535–1539 (2008).
16. Beerli, R. R., Segal, D. J., Dreier, B. & Barbas, C. F. III. Toward controlling gene expression at will: specific regulation of the *erbB-2/HER-2* promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl Acad. Sci. USA* **95,** 14628–14633 (1998).
17. Cong, L., Zhou, R., Kuo, Y.-c., Cunniff, M. & Zhang, F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nature Commun.* **3,** 968 (2012).
18. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).
19. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* http://dx.doi.org/10.1093/nar/gkt520 (2013).
20. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152,** 1173–1183 (2013).
21. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
22. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA* **109,** E2579–E2586 (2012).
23. Banerjee, R. *et al.* The signaling state of *Arabidopsis* cryptochrome 2 contains flavin semiquinone. *J. Biol. Chem.* **282,** 14916–14922 (2007).
24. Moore, M. J. & Proudfoot, N. J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136,** 688–700 (2009).
25. Proudfoot, N. J., Furger, A. & Dye, M. J. Integrating mRNA processing with transcription. *Cell* **108,** 501–512 (2002).
26. Liang, F.-S., Ho, W. Q. & Crabtree, G. R. Engineering the ABA plant stress pathway for regulation of induced proximity. *Sci. Signal.* **4,** rs2 (2011).
27. Holkers, M. *et al.* Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res.* **41,** e63 (2013).
28. Zhang, F. *et al.* Optogenetic interrogation of neural circuits: technology for probing mammalian brain structures. *Nature Protocols* **5,** 439–456 (2010).
29. de Groote, M. L., Verschure, P. J. & Rots, M. G. Epigenetic editing: targeted rewriting of epigenetic marks to modulate expression of selected target genes. *Nucleic Acids Res.* **40,** 10596–10613 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Jennings for comments, F. A. Ran for help with illustrations, C. Lin for editing, M. M. Cunniff for technical assistance and W. Yan for computational analysis, and members of the Zhang laboratory for discussion, support and advice. S.K. is supported by a Hubert Schoemaker Fellowship from the McGovern Institute for Brain Research at MIT. M.H. is supported by a postdoctoral fellowship from

the Human Frontiers Science Program. G.M.C. is support by a NIH NHGRI CEGS grant (P50-HG005550). F.Z. is supported by a NIH Transformative R01 award (R01-NS073124), a NIH Director's Pioneer Award (DP1-MH100706), the Keck, McKnight, Vallee, Damon Runyon, Searle Scholars, Klingenstein, and Simons Foundations, Bob Metcalfe and Jane Pauley. Sequence, protocol, and reagent information are available through the Zhang laboratory website at http://www.genome-engineering.org.

Author Contributions S.K., M.D.B. and F.Z. developed the concept and designed experiments. S.K., M.D.B., A.E.T., P.D.H., M.H. and D.A.S. carried out LITE-related experiments and analysed data. L.C. and P.D.H. developed the SID4X effector domain, the P11-targeting TALEs and the abscisic acid induction system. R.J.P. developed the Cas9 transcription activator and repressor systems. S.K., A.E.T., M.D.B., P.D.H. and F.Z. wrote the manuscript with input from M.H., L.C. and G.M.C.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.Z. (zhang@broadinstitute.org).

METHODS

Design and construction of LITEs. All LITE construct sequences can be found in the Sequences section of the Supplementary Information. We evaluated full-length CRY2 as well as a truncation consisting of the photolyase homology region alone (CRY2PHR, amino acids 1–498)[6]. For CIB1, we tested the full-length protein as well as an amino-terminal domain-only fragment (CIBN, amino acids 1–170)[6]. The efficacy of each design is determined based on the level of light-dependent upregulation of the endogenous target *Neurog2* mRNA (Fig. 1b). To use AAV as a vector for the delivery of LITE components, we needed to ensure that the total viral genome size of each recombinant AAV, with the LITE transgenes included, did not exceed the packaging limit of 4.8 kilobases[30]. We shortened the TALE N and carboxy termini (keeping 136 amino acids in the N terminus and 63 in the C terminus) and exchanged the CRY2PHR (1.5 kb) and CIB1 (1 kb) domains (TALE–CIB1 and CRY2PHR–VP64; Fig. 3a). TALE binding sequences were selected based on DNase I-sensitive regions in the promoter of each target gene. TALE targeting sequences are listed in Supplementary Table 6.

Neuro 2a culture and experiments. Neuro 2a cells (Sigma-Aldrich) were grown in media containing a 1:1 ratio of OptiMEM (Life Technologies) to high-glucose DMEM with GlutaMax and sodium pyruvate (Life Technologies) supplemented with 5% HyClone heat-inactivated FBS (Thermo Scientific), 1% penicillin/streptomycin (Life Technologies), and passaged at 1:5 every 2 days. 120,000 cells were plated in each well of a 24-well plate 18–20 h before transfection. 1 h before transfection, media was changed to DMEM supplemented with 5% HyClone heat-inactivated FBS and 1% penicillin/streptomycin. Cells were transfected with 1.0 μg total of construct DNA (at equimolar ratios) per well with 1.5 μl of GenJet (SignaGen Laboratories) transfection reagent according to the manufacturer's instructions. Media was exchanged 24 h and 44 h post-transfection and light stimulation was started at 48 h. Stimulation parameters were: 5 mW cm^{-2}, 466 nm, 7% duty cycle (1 s light pulse 0.067 Hz) for 12 h unless indicated otherwise in figure legends. RNA was extracted using the RNeasy kit (Qiagen) or NucleoSpin RNA kit (Macherey-Nagel) according to manufacturer's instructions and 1 μg of RNA per sample was reverse-transcribed using qScript (Quanta Biosystems). Relative mRNA levels were measured by reverse transcription and quantitative PCR (qRT–PCR) using TaqMan probes specific for the targeted gene as well as GAPDH as an endogenous control (Life Technologies, see Supplementary Table 7 for TaqMan probe IDs). $\Delta\Delta C_t$ analysis was used to obtain fold-changes relative to negative controls transduced with GFP only and subjected to light stimulation. Toxicity experiments were conducted using the LIVE/DEAD assay kit (Life Technologies) according to manufacturer's protocol.

AAV vector production. The ssDNA-based genome of AAV is less susceptible to recombination, thus providing an advantage over RNA-based lentiviral vectors[27] for the packaging and delivery of highly repetitive TALE sequences. 293FT cells (Life Technologies) were grown in antibiotic-free D10 media (DMEM high glucose with GlutaMax and sodium pyruvate, 10% heat-inactivated HyClone FBS, and 1% 1 M HEPES) and passaged daily at 1:2–2.5. The total number of passages was kept below 10 and cells were never grown beyond 85% confluence. The day before transfection, 10^7 cells in 21.5 ml of D10 media were plated onto 15-cm dishes and incubated for 18–22 h or until ~80% confluence. For use as a transfection reagent, 1 mg ml^{-1} of PEI "Max" (Polysciences) was dissolved in water and the pH of the solution was adjusted to 7.1. For AAV production, 10.4 μg of pDF6 helper plasmid, 8.7 μg of pAAV1 serotype packaging vector, and 5.2 μg of pAAV vector carrying the gene of interest were added to 434 μl of serum-free DMEM and 130 μl of PEI "Max" solution was added to the DMEM-diluted DNA mixture. The DNA/DMEM/PEI cocktail was vortexed and incubated at room temperature for 15 min. After incubation, the transfection mixture was added to 22 ml of complete media, vortexed briefly, and used to replace the media for a 15-cm dish of 293FT cells. For supernatant production, transfection supernatant was collected at 48 h, filtered through a 0.45 μm PVDF filter (Millipore), distributed into aliquots, and frozen for storage at −80 °C.

Primary cortical neuron culture. Dissociated cortical neurons were prepared from C57BL/6N mouse embryos on E16 (Charles River Labs). Cortical tissue was dissected in ice-cold HBSS (50 ml 10× HBSS, 435 ml dH_2O, 0.3 M HEPES pH 7.3, and 1% penicillin/streptomycin). Cortical tissue was washed 3 times with 20 ml of ice-cold HBSS and then digested at 37 °C for 20 min in 8 ml of HBSS with 240 μl of 2.5% trypsin (Life Technologies). Cortices were then washed 3 times with 20 ml of warm HBSS containing 1 ml FBS. Cortices were gently triturated in 2 ml of HBSS and plated at 150,000 cells per well in poly-D-lysine-coated 24-well plates (BD Biosciences). Neurons were maintained in Neurobasal media (Life Technologies), supplemented with 1× B27 (Life Technologies), GlutaMax (Life Technologies) and 1% penicillin/streptomycin.

Primary neuron transduction and light stimulation experiments. Primary cortical neurons were transduced with 250 μl of AAV1 supernatant on DIV 5 (DIV, days *in vitro*). The media and supernatant were replaced with regular complete Neurobasal the following day. Neurobasal was exchanged with Minimal Essential Medium (Life Technologies) containing 1× B27, GlutaMax (Life Technologies) and 1% penicillin/streptomycin 6 days after AAV transduction to prevent formation of phototoxic products from HEPES and riboflavin contained in Neurobasal during light stimulation. For co-transduction of primary neurons with two AAV vectors, the co-delivery efficiency is >80%, with individual components having transduction efficiencies between 83–92%.

Light stimulation was started 6 days after AAV transduction (DIV 11) with an intensity of 5 mW cm^{-2}, duty cycle of 0.8% (250 ms pulses at 0.033 Hz or 500 ms pulses at 0.016 Hz), 466 nm blue light for 24 h unless indicated otherwise in figure legends. RNA extraction and reverse transcription were performed using the Cells-to-Ct kit according to the manufacturer's instructions (Life Technologies). Relative mRNA levels were measured by reverse transcription and quantitative PCR (qRT–PCR) using TaqMan probes as described above for Neuro 2a cells.

Immunohistochemistry of primary neurons. For immunohistochemistry of primary neurons, cells were plated on poly-D-lysine/laminin coated coverslips (BD Biosciences) after harvesting. AAV1-transductions were performed as described above. Neurons were fixed 7 days post-transduction with 4% paraformaldehyde (Sigma Aldrich) for 15 min at RT. Blocking and permeabilization were performed with 10% normal goat serum (Life Technologies) and 0.5% Triton-X100 (Sigma-Aldrich) in DPBS (Life Technologies) for 1 h at room temperature. Neurons were incubated with primary antibodies overnight at 4 °C, washed 3× with DPBS and incubated with secondary antibodies for 90 min at room temperature. For antibody providers and concentrations used, see Supplementary Table 8. Coverslips were finally mounted using Prolong Gold Antifade Reagent with DAPI (Life Technologies) and imaged on an Axio Scope A.1 (Zeiss) with an X-Cite 120Q light source (Lumen Dynamics). Images were acquired using an AxioCam MRm camera and AxioVision 4.8.2.

Western blots. For preparation of total protein lysates, primary cortical neurons were collected after light stimulation (see above) in ice-cold lysis buffer (RIPA, Cell Signaling; 0.1% SDS, Sigma-Aldrich; and cOmplete ULTRA protease inhibitor mix, Roche Applied Science). Cell lysates were sonicated for 5 min at 'M' setting with the Bioruptor water bath sonicator (Diagenode) and centrifuged at 21,000*g* for 10 min at 4 °C. Protein concentration was determined using the RC DC protein assay (Bio-Rad). 30–40 μg of total protein per lane was separated under non-reducing conditions on 4–15% Tris-HCl gels (Bio-Rad) along with Precision Plus Protein Dual Color Standard (Bio-Rad) After wet electrotransfer to polyvinylidene difluoride membranes (Millipore) and membrane blocking for 45 min in 5% BLOT-QuickBlocker (Millipore) in Tris-buffered saline (TBS, Bio-Rad), western blots were probed with anti-mGluR2 (Abcam, 1:1,000) and anti-α-tubulin (Sigma-Aldrich 1:20,000) overnight at 4 °C, followed by washing and anti-mouse-IgG HRP antibody incubation (Sigma-Aldrich, 1:5,000–1:10,000). For further antibody details see Supplementary Table 8. Detection was performed via ECL western blot substrate (SuperSignal West Femto Kit, Thermo Scientific). Blots were imaged with an AlphaImager system (Innotech), and quantified using ImageJ software 1.46r.

Production of concentrated and purified AAV1/2 vectors. Production of concentrated and purified AAV for stereotactic injection *in vivo* was performed using the same initial steps outlined above for production of AAV1 supernatant. However, for transfection, equal ratios of AAV1 and AAV2 serotype plasmids were used instead of AAV1 alone. Five 15-cm plates were transfected per construct and cells were collected with a cell-scraper 48 h post transfection. Purification of AAV1/2 particles was performed using HiTrap heparin affinity columns (GE Healthcare)[31]. We added a second concentration step down to a final volume of 100 μl per construct using an Amicon 500 μl concentration column (100 kDa cutoff, Millipore) to achieve higher viral titres. Titration of AAV was performed by qRT–PCR using a custom TaqMan probe for WPRE (woodchuck hepatitis post-transcriptional response element; Life Technologies). Prior to qRT–PCR, concentrated AAV was treated with DNase I (New England Biolabs) to achieve a measurement of DNase I-resistant particles only. Following DNase I heat-inactivation, the viral envelope was degraded by proteinase K digestion (New England Biolabs). Viral titre was calculated based on a standard curve with known WPRE copy numbers.

Stereotactic injection of AAV1/2 and optical implant. All animal procedures were approved by the MIT Committee on Animal Care. Adult (10–14 weeks old) male C57BL/6N mice were anaesthetized by intraperitoneal (i.p.) injection of ketamine/xylazine (100 mg kg^{-1} ketamine and 10 mg kg^{-1} xylazine) and pre-emptive analgesia was applied (Buprenex, 1 mg kg^{-1}, i.p.). Craniotomy was performed according to approved procedures and 1 μl of AAV1/2 was injected into ILC at 0.35/1.94/−2.94 (lateral, anterior and inferior coordinates in mm relative to bregma). During the same surgical procedure, an optical cannula with fibre (Doric Lenses) was implanted into ILC unilaterally with the end of the optical fibre located at 0.35/1.94/−2.74 relative to bregma. The cannula was affixed to the skull using Metabond dental cement (Parkell Inc.) and Jet denture repair (Lang Dental) to build a

stable, supporting cone. The incision was sutured and proper post-operative analgesics were administered for 3 days following surgery.

Immunohistochemistry on ILC brain sections. Mice were injected with a lethal dose of ketamine/xylazine anaesthetic and transcardially perfused with PBS and 4% paraformaldehyde (PFA). Brains were additionally fixed in 4% PFA at 4 °C overnight and then transferred to 30% sucrose for cryoprotection overnight at room temperature. Brains were then transferred into Tissue-Tek Optimal Cutting Temperature (OCT) Compound (Sakura Finetek) and frozen at −80 °C. 18-μm sections were cut on a cryostat (Leica Biosystems) and mounted on Superfrost Plus glass slides (Thermo Fischer). Sections were post-fixed with 4% PFA for 15 min, and immunohistochemistry was performed as described for primary neurons above.

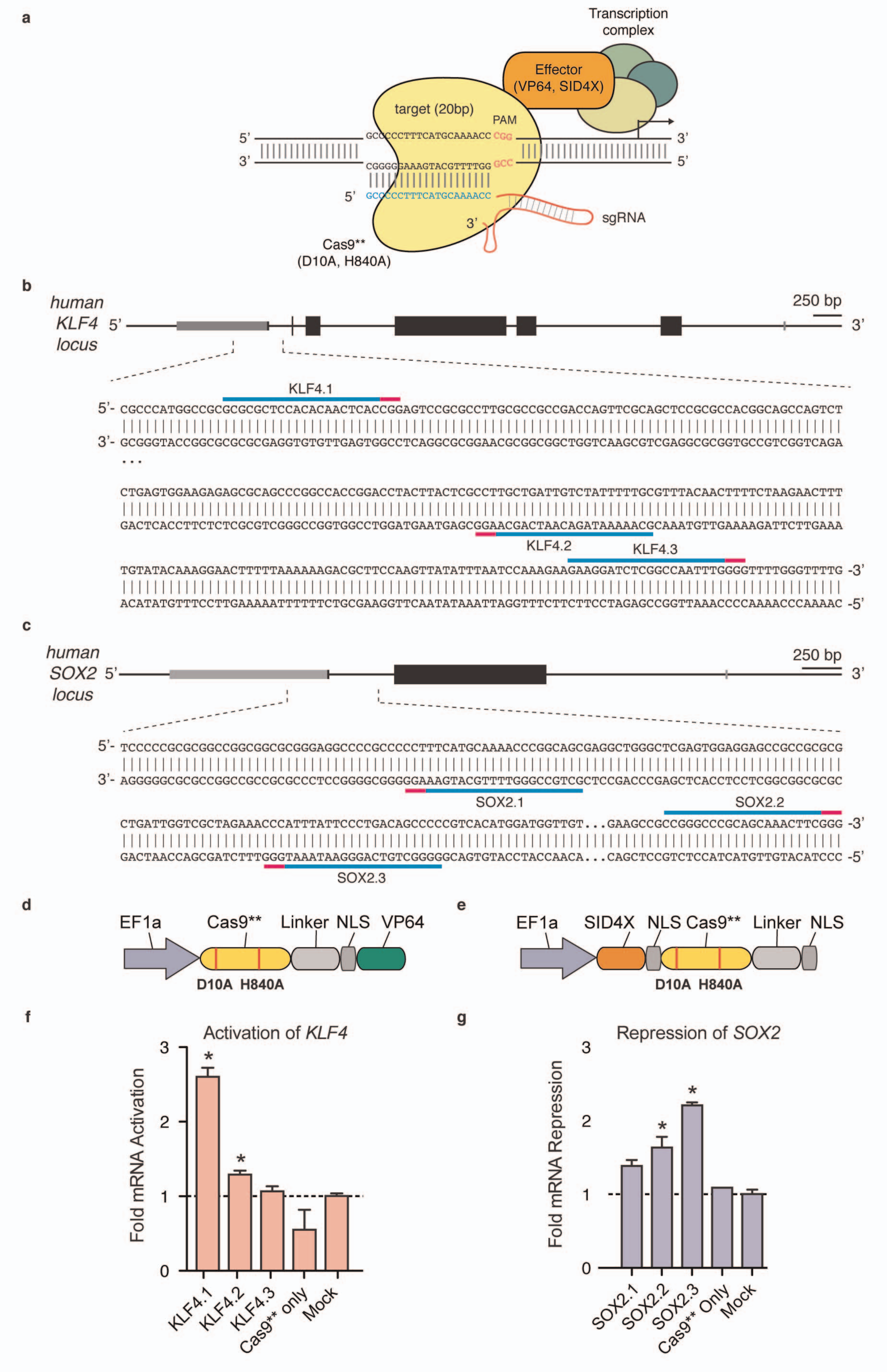
Light stimulation and mRNA level analysis in ILC. Neurons at the injection site were efficiently co-transduced by both viruses, with >80% of transduced cells expressing both TALE(*Grm2*)–CIB1 and CRY2PHR–VP64 (Fig. 3g and Extended Data Fig. 5b). 8 days post-surgery, awake and freely moving mice were stimulated using a 473 nm laser source (OEM Laser Systems) connected to the optical implant via fibre patch cables and a rotary joint. Stimulation parameters were identical to those used on primary neurons: 5 mW (total output), 0.8% duty cycle (500 ms light pulses at 0.016 Hz) for a total of 12 h. Brain tissue from the fibre optic cannula implantation site was analysed (Fig. 3h) for changes in *Grm2* mRNA. Experimental conditions, including transduced constructs and light stimulation are listed in Supplementary Table 9.

After the end of light stimulations, mice were euthanized using CO_2 and the prefrontal cortices (PFC) were quickly dissected on ice and incubated in RNA later (Qiagen) at 4 °C overnight. 200 μm sections were cut in RNA later at 4 °C on a vibratome (Leica Biosystems). Sections were then frozen on a glass cover slide on dry ice and virally transduced ILC was identified under a fluorescent stereomicroscope (Leica M165 FC). A 0.35-mm diameter punch of ILC, located directly ventrally to the termination of the optical fibre tract, was extracted (Harris uni-core, Ted Pella). The brain punch sample was then homogenized using an RNase-free pellet-pestle grinder (Kimble Chase) in 50 μl Cells-to-Ct RNA lysis buffer and RNA extraction, reverse transcription and qRT–PCR was performed as described for primary neuron samples.

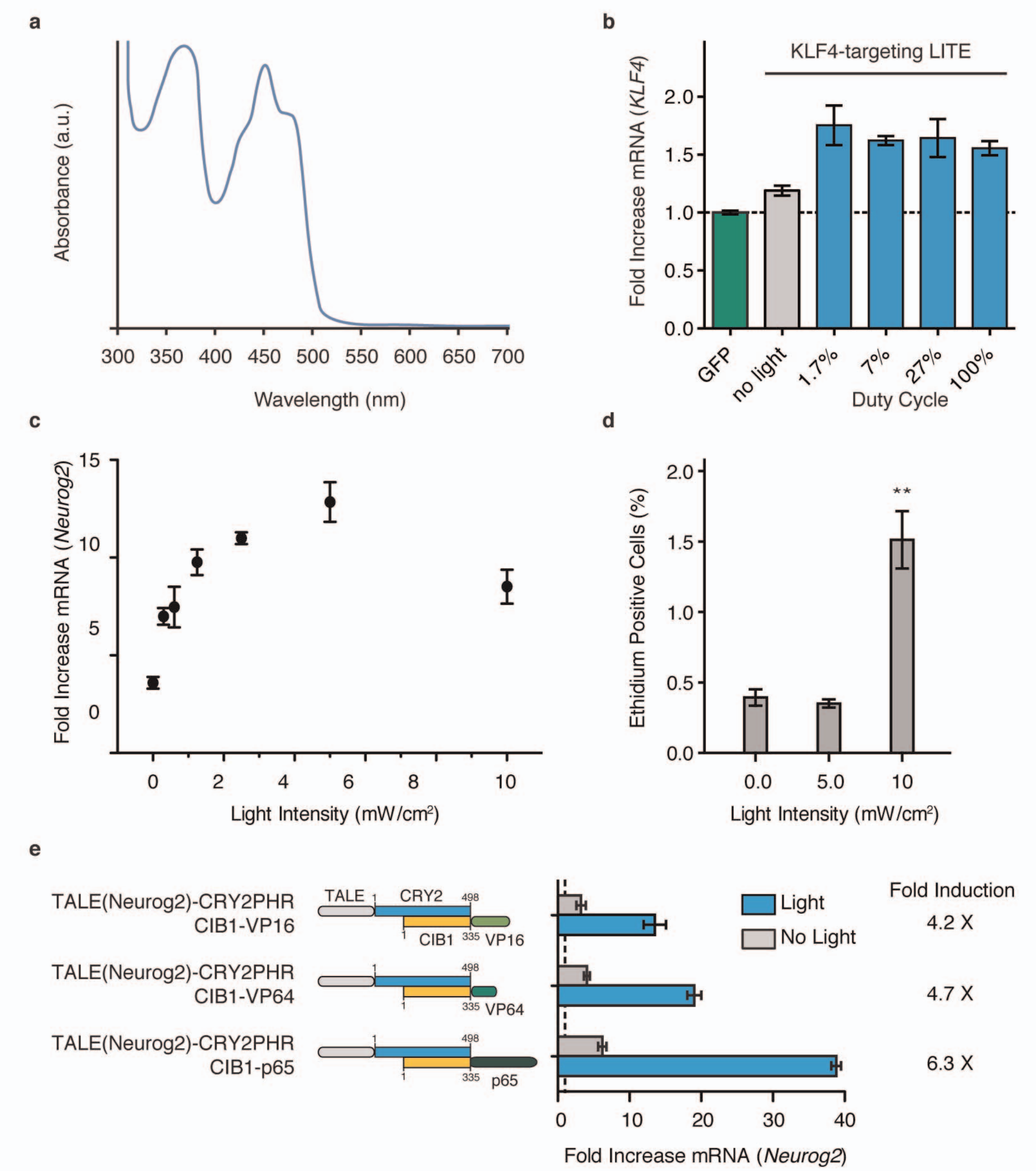
Chromatin immunoprecipitation (ChIP). Neurons or Neuro 2a cells were cultured and transduced or transfected as described above. ChIP samples were prepared as previously described[32] with minor adjustments for the cell number and cell type. Cells were collected in 24-well format, washed in 96-well format, and transferred to microcentrifuge tubes for lysis. Sample cells were directly lysed by water bath sonication with the Biorupter sonication device for 21 min using 30 s on/off cycles and 4 °C chilled circulation (Diagenode). qRT–PCR was used to assess enrichment of histone marks at the targeted locus. qRT–PCR primer sequences are listed in Supplementary Table 10.

Statistical analysis. All experiments were performed with a minimum of two independent biological replicates. Statistical analysis was performed with Prism (GraphPad) using Student's two-tailed *t*-test when comparing two conditions, ANOVA with Tukey's post-hoc analysis when comparing multiple samples with each other, and ANOVA with Dunnett's post-hoc analysis when comparing multiple samples to the negative control.

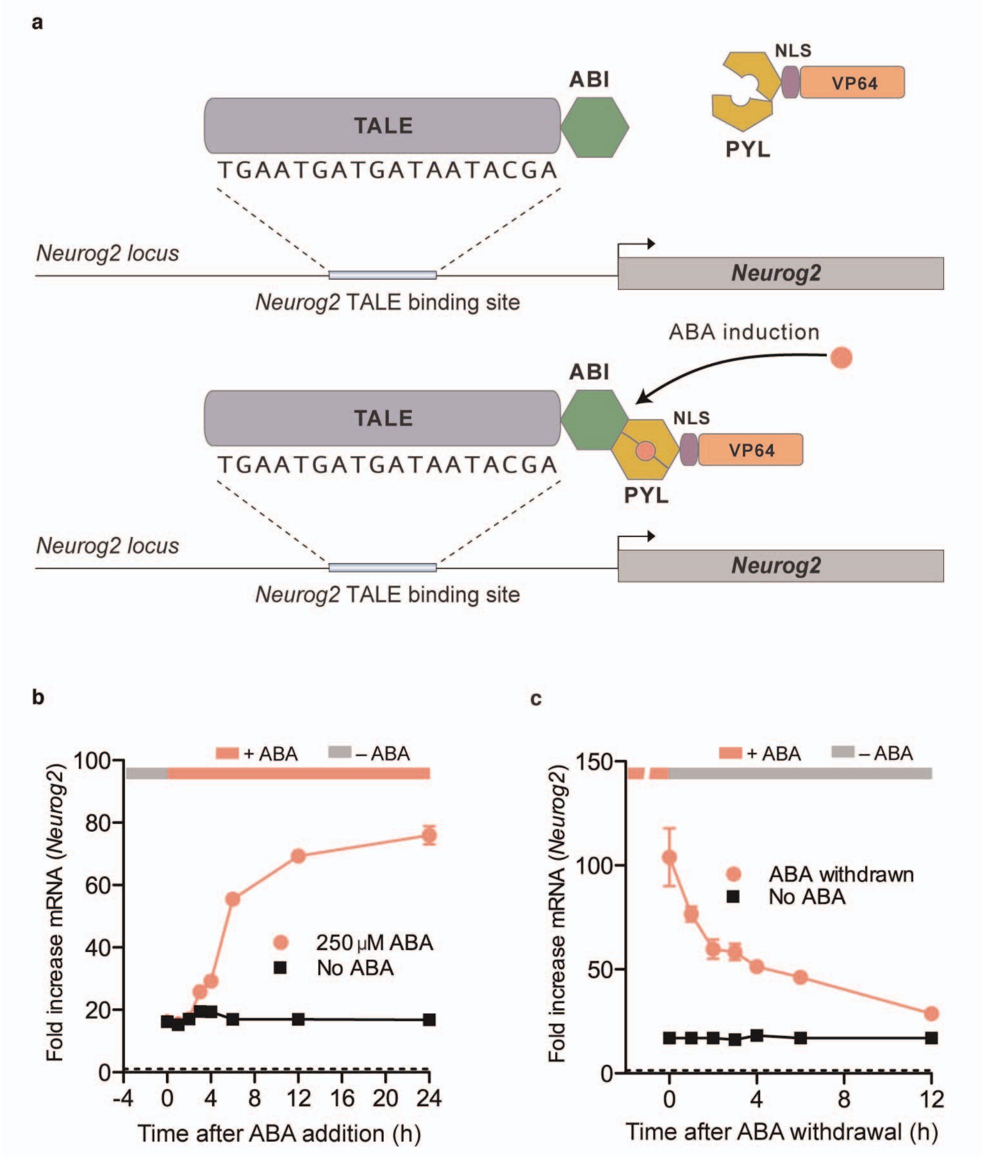
30. Wu, Z., Yang, H. & Colosi, P. Effect of genome size on AAV vector packaging. *Mol. Ther.* **18,** 80–86 (2010).
31. McClure, C., Cole, K. L., Wulff, P., Klugmann, M. & Murray, A. J. Production and titering of recombinant adeno-associated viral vectors. *Vis. Exp.* **57,** e3348 (2011).
32. Blecher-Gonen, R. *et al.* High-throughput chromatin immunoprecipitation for genome-wide mapping of *in vivo* protein-DNA interactions and epigenomic states. *Nature Protocols* **8,** 539–554 (2013).
33. Szymczak, A. L. *et al.* Correction of multi-gene deficiency *in vivo* using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nature Biotechnol.* **22,** 589–594 (2004).
34. Christie, J. M. *et al.* Structural tuning of the fluorescent protein iLOV for improved photostability. *J. Biol. Chem.* **287,** 22295–22304 (2012).


a
Transcription complex
Effector (VP64, SID4X)
target (20bp)
PAM
sgRNA
Cas9** (D10A, H840A)
b
human KLF4 locus
250 bp
KLF4.1
KLF4.2
KLF4.3
c
human SOX2 locus
250 bp
SOX2.1
SOX2.2
SOX2.3
d
EF1a
Cas9**
Linker
NLS
VP64
D10A H840A
e
EF1a
SID4X
NLS
Cas9**
Linker
NLS
D10A H840A
f
Activation of KLF4
Fold mRNA Activation
KLF4.1
KLF4.2
KLF4.3
Cas9** only
Mock
g
Repression of SOX2
Fold mRNA Repression
SOX2.1
SOX2.2
SOX2.3
Cas9** Only
Mock

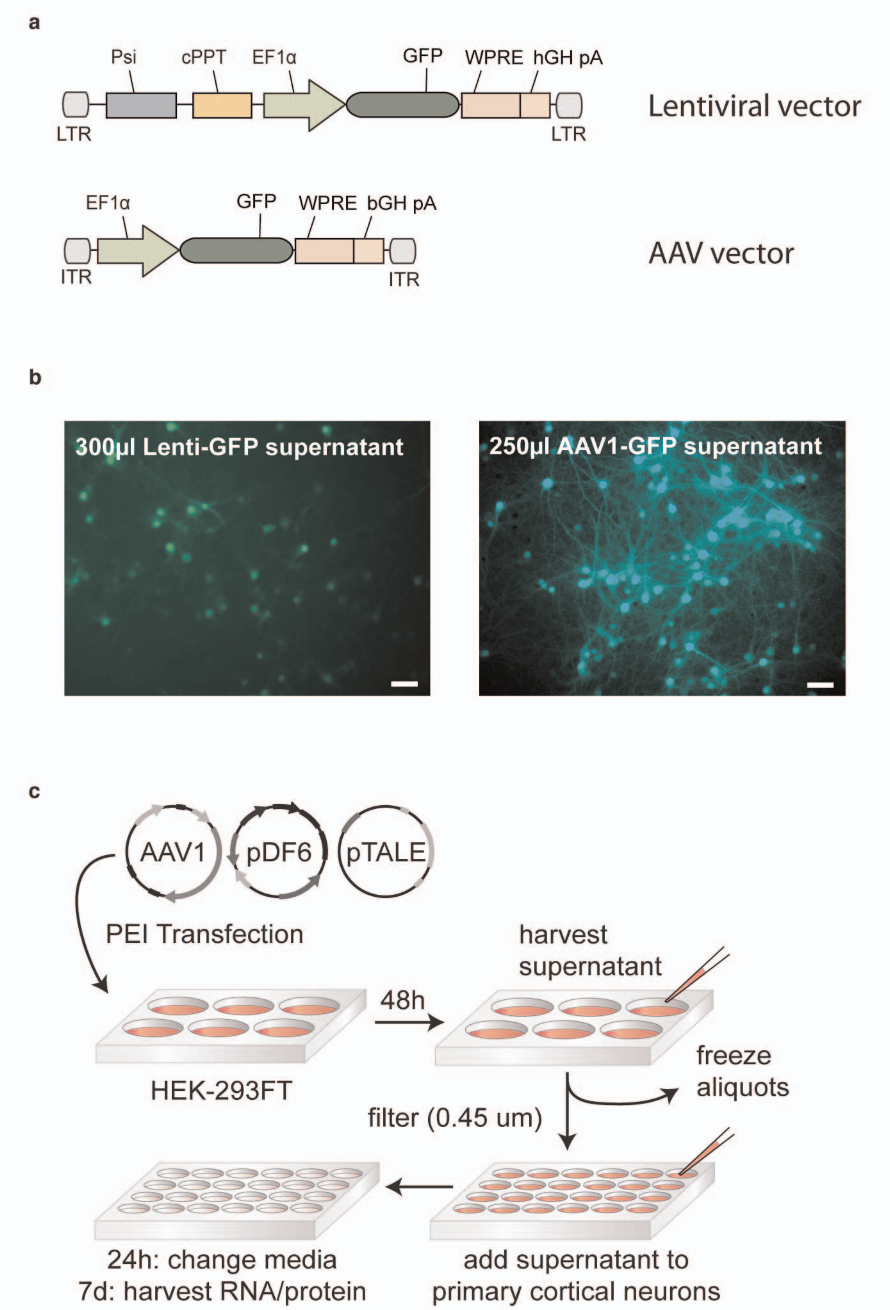
Extended Data Figure 1 | RNA-guided DNA binding protein Cas9 can be used to target transcription effector domains to specific genomic loci. **a**, The RNA-guided nuclease Cas9 from the type II *Streptococcus pyogenes* CRISPR/Cas system can be converted into a nucleolytically inactive RNA-guided DNA binding protein (Cas9**) by introducing two alanine substitutions (D10A and H840A). Schematic showing that a synthetic guide RNA (sgRNA) can direct Cas9**-effector fusion to a specific locus in the human genome. The sgRNA contains a 20-bp guide sequence at the 5′ end which specifies the target sequence. On the target genomic DNA, the 20-bp target site needs to be followed by a 5′-NGG PAM motif. **b**, **c**, Schematics showing the sgRNA target sites in the human *KLF4* and *SOX2* loci, respectively. Each target site is indicated by the blue bar and the corresponding PAM sequence is indicated by the magenta bar. **d**, **e**, Schematics of the Cas9**–VP64 transcription activator and SID4X–Cas9** transcription repressor constructs. **f**, **g**, Cas9**–VP64- and SID4X–Cas9**-mediated activation of *KLF4* and repression of *SOX2*, respectively. All mRNA levels were measured relative to GFP mock-transfected 293FT cells (mean ± s.e.m.; $n = 3$ biological replicates).



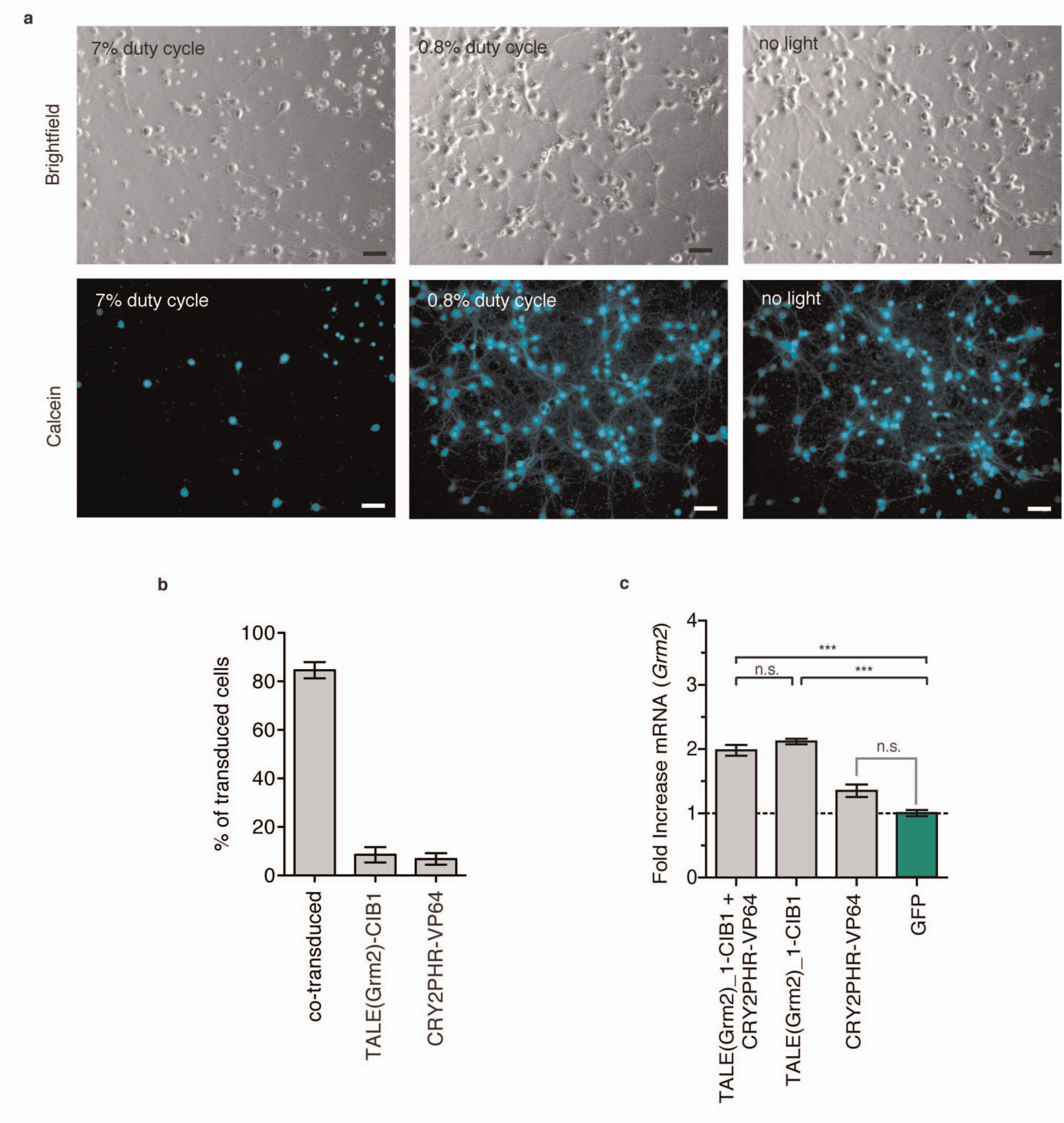
Extended Data Figure 2 | Engineering of light stimulation parameters and activation domains of LITEs. a, Illustration of the absorption spectrum of CRY2 *in vitro*. Cryptochrome 2 was optimally activated by 350–475 nm light[23]. A sharp drop in absorption and activation was seen for wavelengths greater than 480 nm. Spectrum was adapted from ref. 23. **b**, Impact of illumination duty cycle on LITE-mediated gene expression. Varying duty cycles (illumination as percentage of total time) were used to stimulate 293FT cells expressing LITEs targeting the *KLF4* gene, to investigate the effect of duty cycle on LITE activity. *KLF4* expression levels were compared to cells expressing GFP only. Stimulation parameters were: 466 nm, 5 mW cm^{-2} for 24 h. Pulses were performed at 0.067 Hz with the following durations: 1.7% = 0.25 s pulse, 7% = 1 s pulse, 27% = 4 s pulse, 100% = constant illumination. (mean ± s.e.m.; $n = 3$–4 biological replicates.) **c**, The transcriptional activity of CRY2PHR/ CIB1 LITE was found to vary according to the intensity of 466 nm blue light. Neuro 2a cells were stimulated for 12 h at a 7% duty cycle (1 s pulses at 0.067 Hz). All *Neurog2* mRNA levels were measured relative to cells expressing GFP only (mean ± s.e.m.; $n = 3$–4 biological replicates). **d**, Light-induced toxicity measured as the percentage of cells positive for red-fluorescent ethidium homodimer-1 versus calcein-positive cells (mean ± s.e.m.; $n = 3$ biological replicates; $**P < 0.01$). **e**, We compared the activation domains VP16 and p65 in addition to VP64 to test the modularity of the LITE CIB1–effector component. *Neurog2* upregulation with and without light by LITEs using different transcriptional activation domains (VP16, VP64 and p65). Neuro 2a cells transfected with LITE were stimulated for 24 h with 466 nm light at an intensity of 5mW cm^{-2} and a duty cycle of 7% (1 s pulses at 0.067 Hz). All three domains produced a significant light-dependent *Neurog2* mRNA upregulation ($P < 0.001$). We selected VP64 for subsequent experiments due to its lower basal activity in the absence of light-stimulation (mean ± s.e.m.; $n = 3$–4 biological replicates).



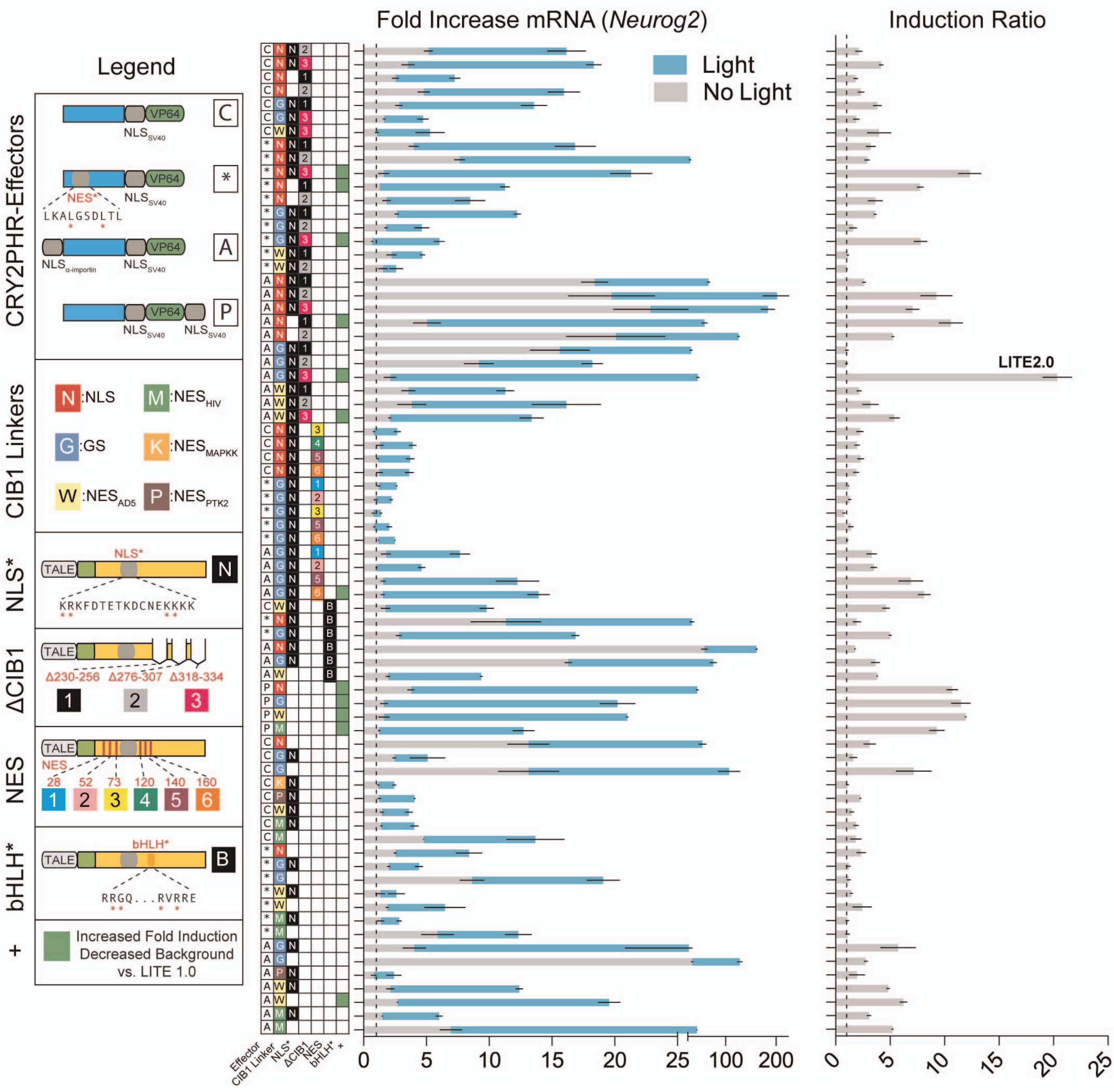
Extended Data Figure 3 | Chemical induction of endogenous gene transcription. **a**, Schematic showing the design of a chemical inducible two-hybrid TALE system based on the abscisic acid (ABA) receptor system. ABI and PYL dimerize upon the addition of ABA and dissociate when ABA is withdrawn. **b**, Time-course of ABA-dependent *Neurog2* upregulation. 250 μM of ABA was added to Neuro 2a cells expressing TALE(*Neurog2*)–ABI and PYL–VP64. Fold mRNA increase was measured at the indicated time points after the addition of ABA. **c**, Decrease of *Neurog2* mRNA levels after 24 h of ABA stimulation. All *Neurog2* mRNA levels were measured relative to GFP-expressing control cells (mean ± s.e.m.; *n* = 3–4 biological replicates).



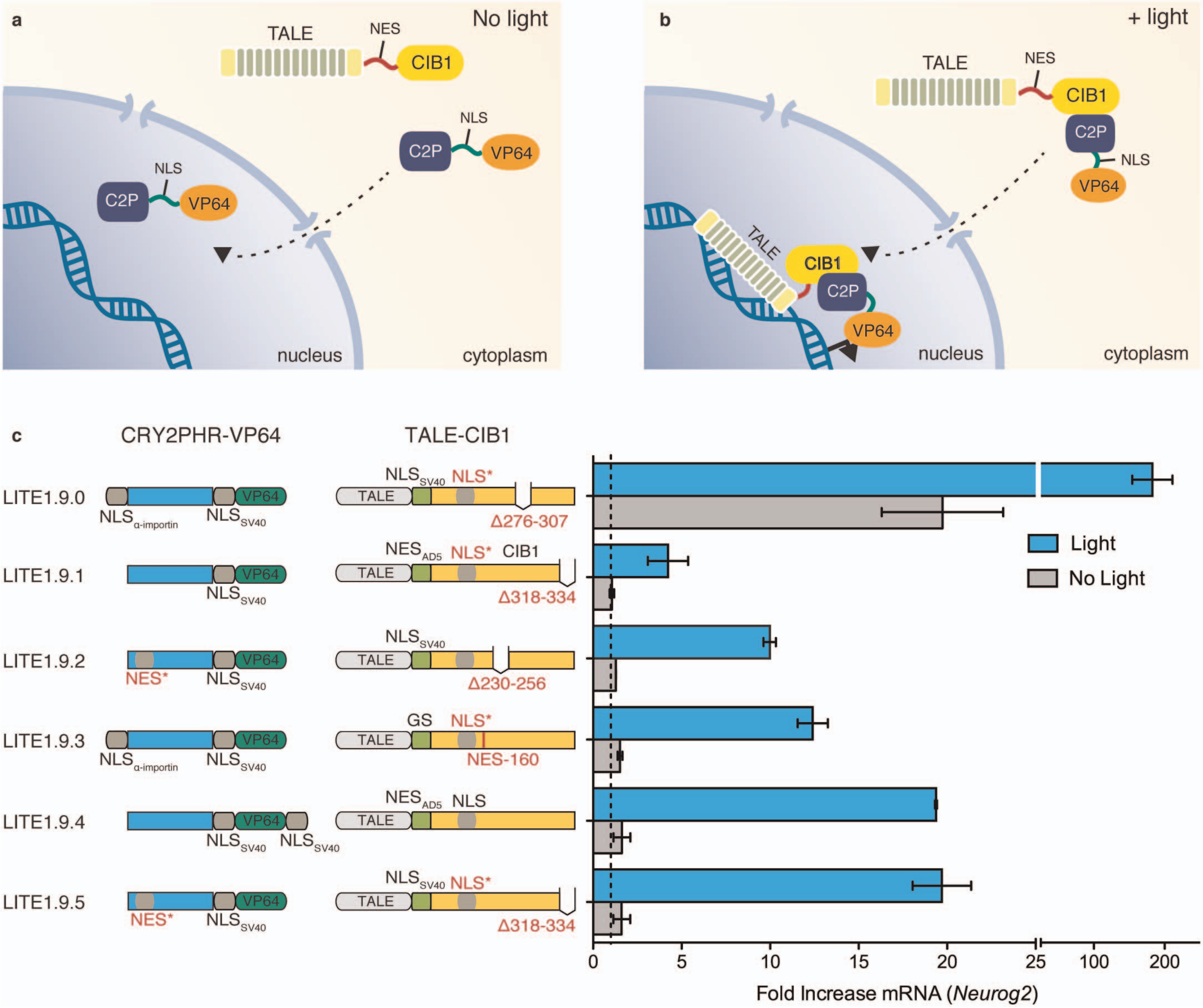
Extended Data Figure 4 | Efficient AAV production using cell supernatant. **a**, Lentiviral and AAV vectors carrying GFP were used to test transduction efficiency. **b**, Primary cortical neurons were transduced with 300 and 250 µl supernatant derived from the same number of lentivirus- or AAV-transduced 293FT cells. Representative images of GFP expression were collected at 7 days post infection. Scale bars, 50 µm. **c**, The depicted process was developed for the production of AAV supernatant and subsequent transduction of primary neurons. 293FT cells were transfected with an AAV vector carrying the gene of interest, the AAV1 serotype packaging vector (pAAV1), and helper plasmid (pDF6) using PEI. 48 h later, the supernatant was collected and filtered through a 0.45 µm PVDF membrane. Primary neurons were then transduced with supernatant and remaining aliquots were stored at −80 °C. Stable levels of AAV construct expression were reached after 5–6 days. AAV supernatant production following this process can be used for production of up to 96 different viral constructs in 96-well format (used for TALE screen in neurons shown in Fig. 2c).



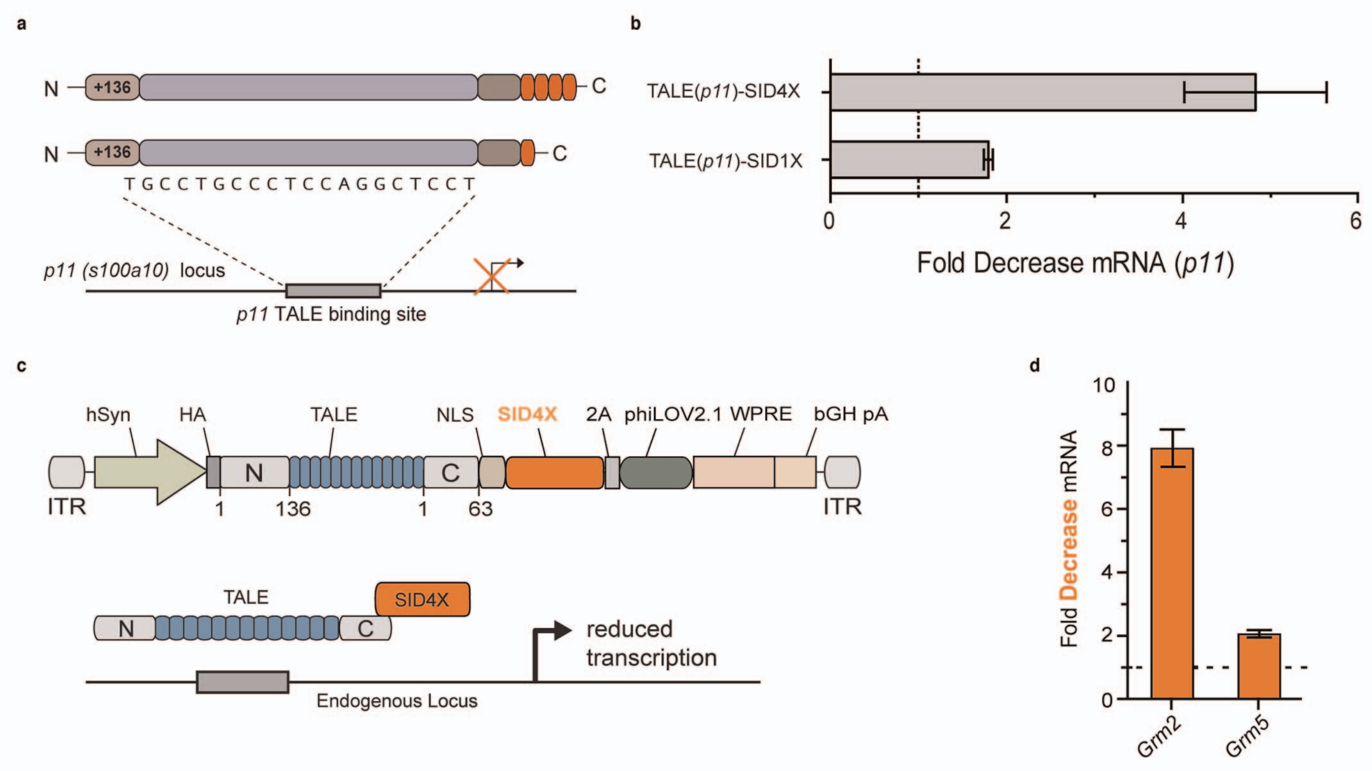
Extended Data Figure 5 | Characterizing LITEs in neurons and *in vivo*. **a**, Impact of light duty cycle on primary neuron health. The effect of light stimulation on primary cortical neuron health was compared for duty cycles of 7%, 0.8%, and no light conditions. Calcein was used to evaluate neuron viability. Bright-field images show cell morphology and integrity. Primary cortical neurons were stimulated with the indicated duty cycle for 24 h with 5 mW cm^{-2} of 466 nm light. Representative images, scale bar, 50 µm. Pulses were performed in the following manner: 7% duty cycle = 1 s pulse at 0.067 Hz, 0.8% duty cycle = 0.5 s pulse at 0.0167 Hz. **b**, Co-transduction efficiency of LITE components by AAV1/2 *in vivo* in mouse infralimbic cortex. Cells transduced by TALE(*Grm2*)–CIB1 alone, CRY2PHR–VP64 alone, or co-transduced were calculated as a percentage of all transduced cells (mean ± s.e.m.; $n = 9$ fields from 3 animals). **c**, *Grm2* mRNA levels were determined in primary neurons transfected with individual LITE components. Primary neurons expressing TALE(*Grm2*)–CIB1 alone led to a similar increase in *Grm2* mRNA levels as unstimulated cells expressing the complete LITE system (mean ± s.e.m.; $n = 3$–4 biological replicates).



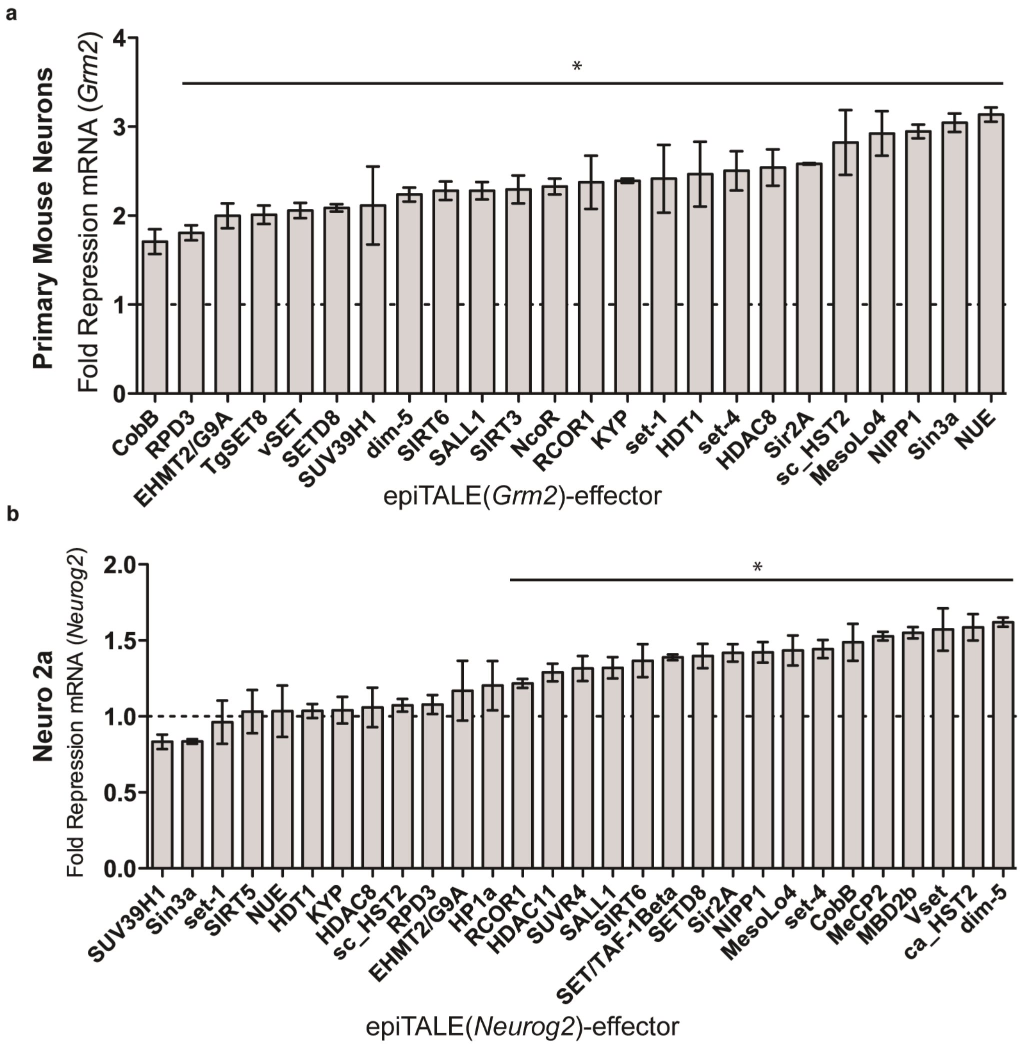
Extended Data Figure 6 | Effects of LITE component engineering on activation, background signal and fold induction. Protein modifications were used to find LITE components resulting in reduced background transcriptional activation while improving induction ratio by light. In brief, nuclear localization signals and mutations in an endogenous nuclear export signal were used to improve nuclear import of the CRY2PHR–VP64 component. Several variations of CIB1 intended to either reduce nuclear localization or CIB1 transcriptional activation were pursued to reduce the contribution of the TALE–CIB1 component to background activity. The results of all tested combinations of CRY2PHR–VP64 and TALE–CIB1 are shown above. The table to the left of the bar graphs indicates the particular combination of domains/mutations used for each condition. Each row of the table and bar graphs contains the component details, light/no light activity, and induction ratio by light for the particular CRY2PHR/CIB1 combination. Combinations that resulted in both decreased background and increased fold induction compared to LITE1.0 are highlighted in green in the table column marked '+' (t-test $P < 0.05$). See Supplementary Discussion for detailed explanation of each modification (mean ± s.e.m.; n = 2–3 biological replicates).



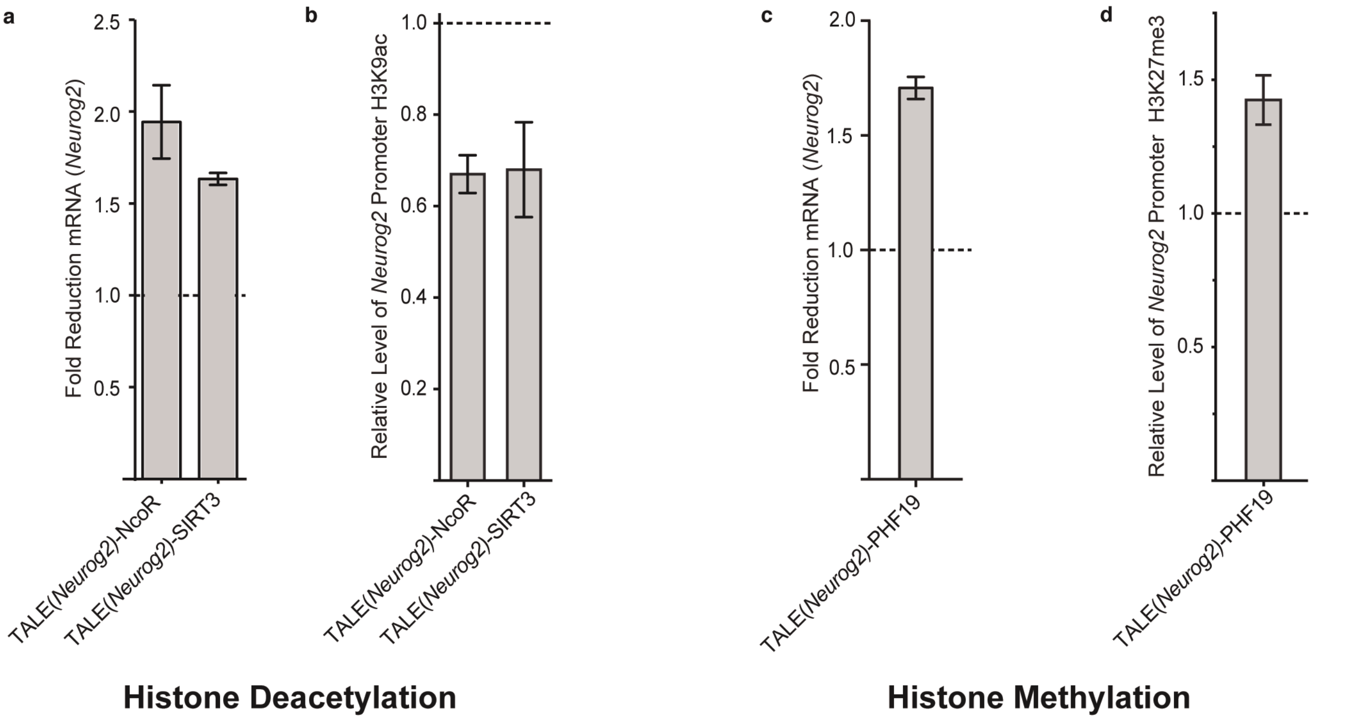
Extended Data Figure 7 | Strategies for optimizing the LITE system. **a**, In the absence of light, the TALE–CIB1 LITE component resides in the cytoplasm due to the absence of a nuclear localization signal, NLS (or the addition of a nuclear export signal, NES). The CRY2PHR–VP64 component containing a NLS on the other hand is actively imported into the nucleus on its own. **b**, In the presence of blue light, TALE–CIB1 binds to CRY2PHR. The NLS present in CRY2PHR–VP64 now mediates nuclear import of the complex of both LITE components, enabling them to activate transcription at the targeted locus. In addition to the LITE2.0 constructs, several CRY2PHR–VP64/TALE–CIB1 combinations from the engineered LITE component screen were of particular note. LITE1.9.0, which combined the α-importin NLS effector construct with a mutated endogenous NLS and Δ276–307 TALE–CIB1 construct, exhibited an induction ratio greater than 9 and an absolute light activation of more than 180. LITE1.9.1, which combined the unmodified CRY2PHR–VP64 with a mutated NLS, Δ318–334, AD5 NES TALE–CIB1 construct, achieved an induction ratio of 4 with a background activation of 1.06. A selection of other LITE1.9 combinations with background activations lower than 2 and induction ratios ranging from 7 to 12 were also highlighted (mean ± s.e.m.; $n = 2$–3 biological replicates).



Extended Data Figure 8 | TALE SID4X repressor characterization and application in neurons. **a**, A synthetic repressor was constructed by concatenating 4 SID domains (SID4X). To identify the optimal TALE-repressor architecture, SID or SID4X was fused to a TALE designed to target the mouse *p11* (also known as *S100a10*) gene. **b**, Fold decrease in *p11* mRNA was assayed using qRT–PCR (mean ± s.e.m.; $n = 3$ biological replicates). **c**, General schematic of constitutive TALE transcriptional repressor packaged into AAV. Effector domain SID4X is highlighted. hSyn, human synapsin promoter; 2A, *Thosea asigna* virus 2A self-cleaving peptide[33]; WPRE, woodchuck hepatitis post-transcriptional response element; bGH pA, bovine growth hormone poly-A signal. phiLOV2.1[34] (330 bp) was chosen as a shorter fluorescent marker to ensure efficient AAV packaging. **d**, A TALE targeting either the endogenous mouse locus *Grm5* or *Grm2* was fused to SID4X and virally transduced into primary neurons. SID4X-mediated target gene downregulation is shown for each TALE relative to levels in control neurons expressing GFP only (mean ± s.e.m.; $n = 3$–4 biological replicates).



Extended Data Figure 9 | A diverse set of epiTALEs mediate transcriptional repression in neurons and Neuro 2a cells. **a**, 24 different histone effector domains were each fused to a *Grm2* targeting TALE. TALE–effector fusions were expressed in primary cortical mouse neurons using AAV transduction. *Grm2* mRNA levels were measured using qRT–PCR relative to neurons transduced with GFP only. (* $P < 0.05$; mean ± s.e.m.; $n = 2$–3 biological replicates.) **b**, A total of 32 epiTALEs were transfected into Neuro2A cells. 20 of them mediated significant repression of the targeted *Neurog2* locus (* $P < 0.05$; mean ± s.e.m.; $n = 2$–3 biological replicates).



Extended Data Figure 10 | epiTALEs mediating transcriptional repression along with histone modifications in Neuro 2a cells. **a**, TALEs fused to histone-deacetylating epigenetic effectors NcoR and SIRT3 targeting the murine *Neurog2* locus in Neuro 2a cells were assayed for repressive activity on *Neurog2* transcript levels (mean ± s.e.m.; $n = 2$–3 biological replicates). **b**, ChIP qRT–PCR showing a reduction in H3K9 acetylation at the *Neurog2* promoter for NcoR and SIRT3 epiTALEs (mean ± s.e.m.; $n = 2$–3 biological replicates). **c**, The epigenetic effector PHF19 with known histone methyltransferase binding activity was fused to a TALE targeting *Neurog2*. Repression of *Neurog2* mRNA levels was observed (mean ± s.e.m.; $n = 2$–3 biological replicates). **d**, ChIP qRT–PCR showing an increase in H3K27me3 levels at the *Neurog2* promoter for the PHF19 epiTALE (mean ± s.e.m.; $n = 2$–3 biological replicates).

doi:10.1038/nature12433

Charting a dynamic DNA methylation landscape of the human genome

Michael J. Ziller[1,2,3], Hongcang Gu[1], Fabian Müller[3]†, Julie Donaghey[1,2,3], Linus T.-Y. Tsai[4], Oliver Kohlbacher[5], Philip L. De Jager[1,6], Evan D. Rosen[1,4], David A. Bennett[7], Bradley E. Bernstein[1,8], Andreas Gnirke[1] & Alexander Meissner[1,2,3]

DNA methylation is a defining feature of mammalian cellular identity and is essential for normal development[1,2]. Most cell types, except germ cells and pre-implantation embryos[3–5], display relatively stable DNA methylation patterns, with 70–80% of all CpGs being methylated[6]. Despite recent advances, we still have a limited understanding of when, where and how many CpGs participate in genomic regulation. Here we report the in-depth analysis of 42 whole-genome bisulphite sequencing data sets across 30 diverse human cell and tissue types. We observe dynamic regulation for only 21.8% of autosomal CpGs within a normal developmental context, most of which are distal to transcription start sites. These dynamic CpGs co-localize with gene regulatory elements, particularly enhancers and transcription-factor-binding sites, which allow identification of key lineage-specific regulators. In addition, differentially methylated regions (DMRs) often contain single nucleotide polymorphisms associated with cell-type-related diseases as determined by genome-wide association studies. The results also highlight the general inefficiency of whole-genome bisulphite sequencing, as 70–80% of the sequencing reads across these data sets provided little or no relevant information about CpG methylation. To demonstrate further the utility of our DMR set, we use it to classify unknown samples and identify representative signature regions that recapitulate major DNA methylation dynamics. In summary, although in theory every CpG can change its methylation state, our results suggest that only a fraction does so as part of coordinated regulatory programs. Therefore, our selected DMRs can serve as a starting point to guide new, more effective reduced representation approaches to capture the most informative fraction of CpGs, as well as further pinpoint putative regulatory elements.

Changes in DNA methylation patterns and the resulting DMRs have been the focus of numerous studies in the context of normal development[7] and disease[8]. These studies have characterized many different DMR classes including partially methylated domains[9], condition-specific[10], cell-type-specific[9,11–13] and tissue-specific DMRs[14,15], as well as DMRs arising in diseases such as cancer[15,16]. Owing to the relatively small fraction of genomic CpGs assayed or small sample cohorts, the question of what fraction of genomic CpGs changes its methylation state in the context of normal development as well as their regulatory context remains underexplored.

In this study, we systematically investigated the DNA methylation state of most human autosomal CpGs to determine those that show dynamic changes and hence may participate in genome regulation in a developmental context (dynamic CpGs). In total, we included 42 whole-genome bisulphite sequencing (WGBS) data sets, comprising a range of human cell and tissue types ($n = 30$). The combined 40.4 billion reads enabled us to assay 25.71 million autosomal CpGs ($\geq 5\times$ coverage in at least $\geq 50\%$ of all samples; 96% of all hg19 autosomal CpGs) (Supplementary Table 1). We organized the samples into four classes: human embryonic stem (ES) cells and human ES-cell-derived cell populations, primary cells, disease conditions, and long-term cultured cell lines (Fig. 1a and Supplementary Table 1). On a global scale, human ES cells and their derivatives exhibit the highest DNA methylation levels, followed by primary cells (~5% less), which is in sharp contrast to the global hypomethylation observed in colon cancer (~10–15% less) and long-term cultured cell lines (10–30% less).

Focusing initially on our developmental sample set ($n = 24$ total, ES cells, *in-vitro*-derived cell types and primary cells; Supplementary Table 1) we identified ~5.6 million dynamic CpGs (minimum methylation difference ≥ 0.3, false discovery rate (FDR) = 10.4%, 21.8% of captured autosomal CpGs; Fig. 1b, Supplementary Fig. 1e and Supplementary Information) distributed across 716,087 discrete DMRs (19.2% of the mappable human genome; Supplementary Table 2). In addition to this moderately stringent cut-off, we also tested thresholds as low as 10% methylation difference that may account for DNA methylation changes arising from relevant small subpopulations in heterogeneous tissue samples or noise, but still only find 10.4 million CpGs to be dynamic (Supplementary Fig. 1a–d).

Focusing on the more stringent set (≥ 0.3 difference), we find approximately 70% are on average highly methylated (>75% methylation ratio), whereas less than 2% are on average unmethylated (<10% methylation ratio) (Supplementary Fig. 1h). In line with this observation, we find that hypomethylation of DMRs shows greater sample specificity than hypermethylation (Fig. 1c). Interestingly, most of the DMRs are small (>75% are smaller than 1 kilobase (kb); Supplementary Fig. 1i) and located distal to transcription start sites (Supplementary Fig. 1j). However, the average variation in DNA methylation levels across all RefSeq promoters ($n = 30{,}090$) does still exhibit a clear increase specifically at the transcription start sites, with most of this variation occurring at intermediate and low CpG density promoters (Fig. 1d). For CpG islands in general, we observe distinct dynamic regimes, highlighting that different classes of CpG islands are probably subject to different modes of regulation[12,17,18] (Fig. 1d, bottom). Consistent with previous reports[15], we find CpG island shores (regions within 2 kb of an island)[15] to be among the most variable genomic regions (Supplementary Fig. 1o). These observations are exemplified at the *OCT4* (also known as *POU5F1*) locus, in which the promoter and large parts of the gene body exhibit high DNA methylation dynamics, whereas the strong downstream CpG island as well as the surrounding CTCF-binding sites remain static (Fig. 1e). Only 12.2% of our DMR set overlap with at least one of four annotated classic, gene-centric genomic features (promoter, exon, CpG island (CGI), or CGI-shore; $n = 568{,}430$) (Fig. 1f). To gain insights into the role of the remaining set, we first investigated their co-localization with DNase I hypersensitive sites across 92 distinct cell types[19] as well as a catalogue of putative enhancer elements for 31 cell and tissue types[20]. Notably, we found that 42.3% of our DMRs overlap with at least one

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [2]Harvard Stem Cell Institute, Cambridge, Massachusetts 02138, USA. [3]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [4]Division of Endocrinology, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. [5]Applied Bioinformatics, Center for Bioinformatics and Quantitative Biology Center, University of Tübingen, 72074 Tübingen, Germany. [6]Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Departments of Neurology and Psychiatry, Brigham and Women's Hospital, 77 Avenue Louis Pasteur, NRB168, Boston, Massachusetts 02115, USA. [7]Rush Alzheimer's Disease Center, Rush University Medical Center, 600 South Paulina Street, Chicago, Illinois 60612, USA. [8]Department of Pathology, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA. †Present address: Max Planck Institute for Informatics, 66123 Saarbrücken, Germany.

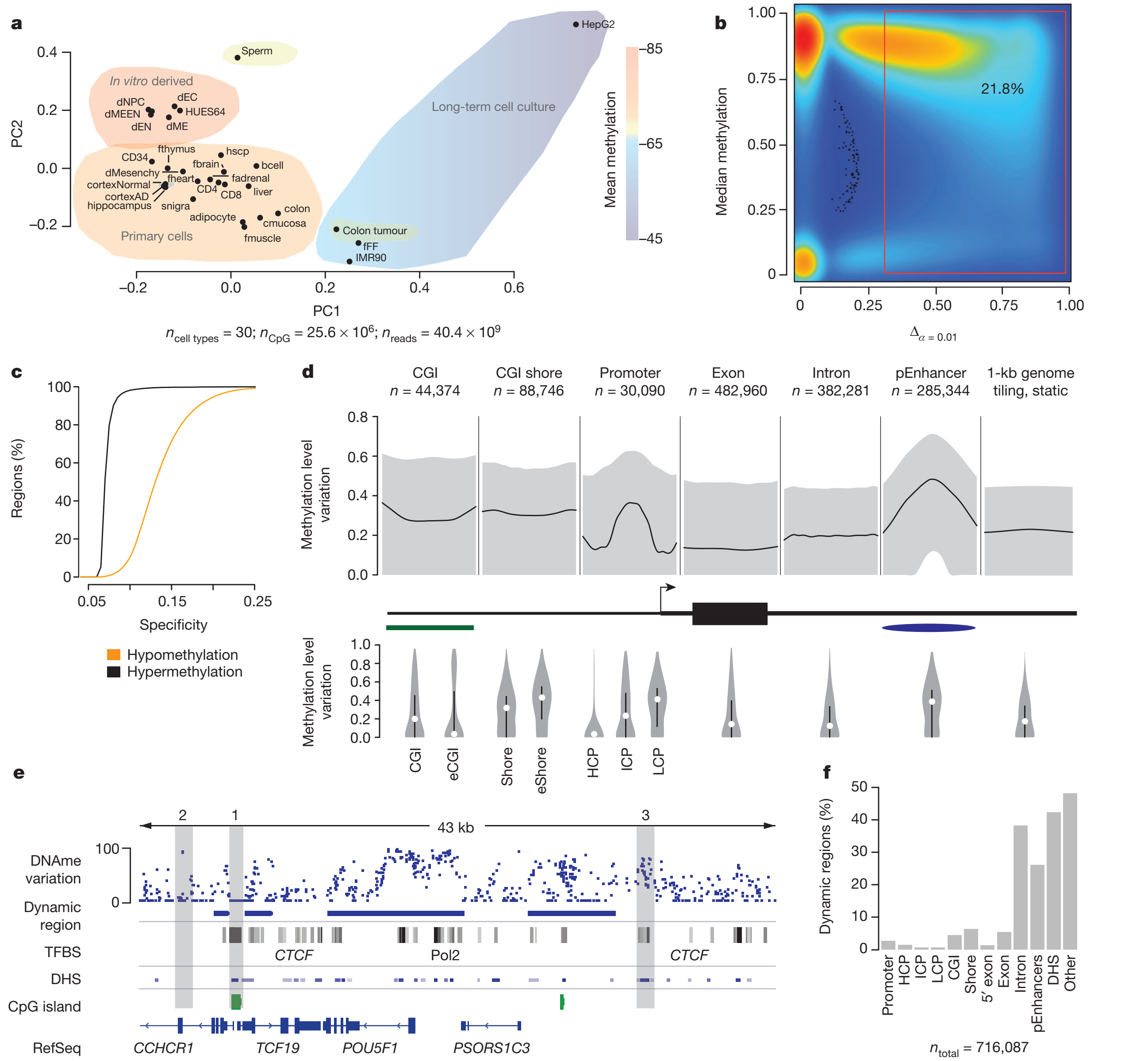


Figure 1 | Identification and characteristics of DMRs in the human genome. **a**, Principal component (PC) analysis based on CpG methylation levels for 1-kb tiles across 30 diverse human cell and tissue samples. Colouring indicates classification of samples into subgroups and group-wise mean DNA methylation. Detailed sample annotations are listed in Supplementary Table 1. Grey area indicates Alzheimer's disease (AD) samples. **b**, Density scatterplot of CpG-wise DNA methylation level differences (x axis, $P \leq 0.01$) and CpG median methylation (y axis) across the 24 developmental samples (excluding cancer and long-term culture). Colouring indicates CpG density from low (blue) to high (red). The red box highlights dynamic CpGs ($\geq$0.3). **c**, Cumulative distribution of DMR specificity. High hypo/hypermethylation specificity indicates that a particular region is methylated/unmethylated in most tissues and deviates from this default state in only one or a few cases. **d**, Top, composite plot of mean DNA methylation differences across various genomic features. Black lines indicate the median of the average DNA methylation difference across each feature. Grey areas mark twenty-fifth and seventy-fifth percentiles. Bottom, distribution of mean DNA methylation difference for each genomic feature. Black bars indicate twenty-fifth and seventy-fifth percentiles; white dots mark the median. For CGI islands, a smaller, experimentally determined set (eCGI; $n = 25{,}490$) is also shown. Promoters are broken down into high CpG content (HCP; $n = 24{,}899$), intermediate CpG content (ICP; $n = 10{,}920$) and low CpG content (LCP; $n = 7{,}946$) regions ($n = 43{,}765$ total). Shore denotes regions within 2 kb of an island; eShore denotes experimentally determined shore. pEnhancer, putative enhancer. **e**, Methylation level variation across the *OCT4* locus (chr6:31,119,000–31,162,000) (top). Blue bars indicate significant DMRs at $P \leq 0.01$, and exhibit a minimum difference ≥ 0.3 across the 24 developmental samples. Grey boxes (1–3) are examples of regions that are static (1 and 2) or that do not meet the threshold of dynamic (3). For reference, ENCODE TFBS cluster track, DNase I hypersensitive sites, CpG islands and RefSeq genes are shown. DNAme, DNA methylation. **f**, Distribution of DMRs across various genomic features. Each region is assigned to only one of these genomic features according to the ranking promoter, CGI, CGI shore, 5′ exon, exon, intron, putative enhancers, DNase I hypersensitive site (DHS) or other.

DNase I hypersensitive site (Fig. 1f), and 26.1% co-localize with enhancer-like regions, which cover more than 50% of all H3K27ac regions in our catalogue ($n = 285{,}344$) and represents one of the most differentially methylated features (Fig. 1d). Next, we examined DMR overlap with transcription-factor-binding site (TFBS) clusters compiled from 165 transcription factors profiled by the ENCODE project[21] and uncovered a highly significant overlap of the two feature classes (odds ratio 1.14, $P < 0.01$ empirical test, Supplementary Information). Interestingly, we find that more than 50% of all DMRs overlap with at least one and 25% with more than three TFBSs, accounting for an additional 13.0% of DMRs (Fig. 2a). Consistent with this, we find markedly increased variation in DNA methylation levels specifically across TFBSs (Supplementary Fig. 2a). In summary, we were able to attribute 64.2% of all DMRs to at least one putative gene regulatory element or coding sequence

(Supplementary Fig. 1e–h), suggesting that they demarcate various classes of regulatory elements.

We determined all cell-type-specific hypomethylated regions ($n = 430{,}250$; see Supplementary Information) and investigated the enrichment for 161 ENCODE factors (excluding MBD4, SETDB1, POL2P and HDAC2 from the previous set). Notably, we observe significant enrichment of cell-type-specific transcription factors that are known to be involved in the regulation of the respective cellular states (Fig. 2b). For instance, the top three factors enriched in HUES64-specific DMRs are OCT4, SOX2 and NANOG (Fig. 2b). Similarly, PU.1 and TAL1 are highly enriched in CD34 cells and hepatocyte nuclear factors in adult liver (Fig. 2b). In further support of this, motif enrichment analysis revealed many more interesting cell-type-specific transcription factor associations, such as enrichment of distinct NKX factors in fetal heart and brain, and ESRRG in fetal adrenal cells (Supplementary Fig. 2b and Supplementary Table 3). Moreover, we tested whether the DMR set can be used to gain insights into the combinatorial control of cellular states by transcription factors. To that end, we determined all unmethylated (<10% methylation) PAX5 motif instances (±100 base pairs (bp)) across the human genome in CD34 or fetal brain cells (Fig. 2c). Although both footprint sets show a large overlap (11,031 sites), regions exclusively unmethylated in CD34 or fetal brain are enriched for distinct sets of other known lineage-specific transcription factor motifs; such as PU.1 in CD34 and LMX1A or EN1 in fetal brain (Fig. 2c). Taken together, these findings highlight that cell-type-specific DNA methylation patterns can be used to detect footprints and infer potential co-regulation by transcription factors. In fact, more than 60% of all ENCODE TFBSs are hypermethylated in most samples, but become hypomethylated very specifically in only one or two cell types (Fig. 2d), whereas 25% are constitutively unmethylated and never change (Fig. 2d).

Breaking down this distribution of TFBSs reveals distinct patterns of variation for different types of transcription factor (Supplementary Fig. 2e). More generally, we find that DNA methylation variation across TFBSs is strongly correlated with its median methylation level and therefore the (hypo-) methylation specificity (Supplementary Fig. 2c), as well as the tissue specificity of transcription factor expression[22] (Supplementary Fig. 2d). These observations support the notion[23] that selective transcription factor binding creates and/or maintains spatially highly constrained hypomethylated regions and confers cell type specificity.

On the basis of these findings and previous reports[24], we asked whether DMRs are more susceptible to point mutations that are functionally consequential. Even with strict filtering criteria, we found a significant enrichment of single nucleotide polymorphisms (SNPs) in DMRs compared to genomic background as well as different sets of random control regions (odds ratio 1.06, $P < 10^{-16}$, binomial test; Supplementary Information). We then determined the overlap of DMRs with recently evolved human-specific CpGs, termed CpG beacons[25], which shows a marked enrichment (odds ratio 1.37–1.6 compared to genomic background and random control regions, $P < 10^{-16}$). This suggests overall higher genetic intra-species variability specifically at regions that change their DNA methylation state. In concordance with the increased SNP frequency, DMRs are also significantly enriched for genome-wide association study (GWAS) SNPs from the GWAS catalogue[26] (odds ratio 1.16, $P = 3.27 \times 10^{-10}$, binomial test). Similar to our observations on TFBSs, GWAS SNPs exhibit a non-random enrichment distribution across cell-type-specific DMRs (Fig. 3). For instance, we find DMRs specific to adult liver to be enriched for liver and serum metabolite-related GWAS SNPs, fetal heart DMRs enriched for cardiovascular-disease SNPs, and many of our blood-cell-type DMRs enriched for autoimmune disease and haematological parameter related SNPs.

It is well known that many cancers exhibit considerable DNA methylation changes[27], we therefore compared a colon cancer to a matched control and found 532,665 differentially methylated CpGs. Forty per cent of these overlapped with the previously identified developmental dynamic set (Fig. 4a). Similarly, 37% of differentially methylated CpGs found in Alzheimer's disease samples compared to normal controls

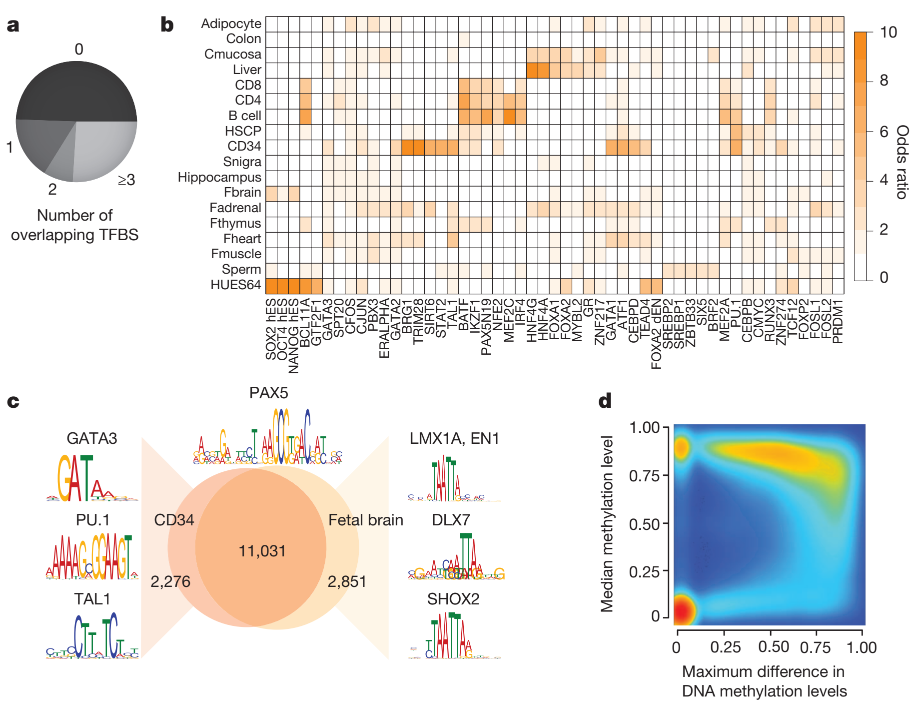


Figure 2 | Dynamic CpG methylation regions frequently co-localize with TFBSs. a, Overlap of DMRs with ENCODE TFBSs. **b**, Enrichment of the top four TFBSs significantly overrepresented ($P < 0.01$, empirical test) in DMRs specific to the cell type indicated (specificity > 0.15). Colour code quantifies median enrichment odds ratio compared to size-matched random control regions. Cmucosa, colon mucosa; F denotes fetal. **c**, Overlap of PAX5 motifs (±100-bp; top) unmethylated in CD34 cells or fetal brain across the entire human genome. Regions specifically unmethylated in CD34 or fetal brain were subjected to motif analysis, and top differentially co-occurring motifs are highlighted on the left for CD34 and on the right for fetal brain. **d**, Density scatterplot of maximum DNA methylation difference across 24 developmental samples for TFBS cluster track ($n = 2.7$ million) and median methylation level across all samples. Colour code indicates density of TFBSs from low (blue) to high (red).

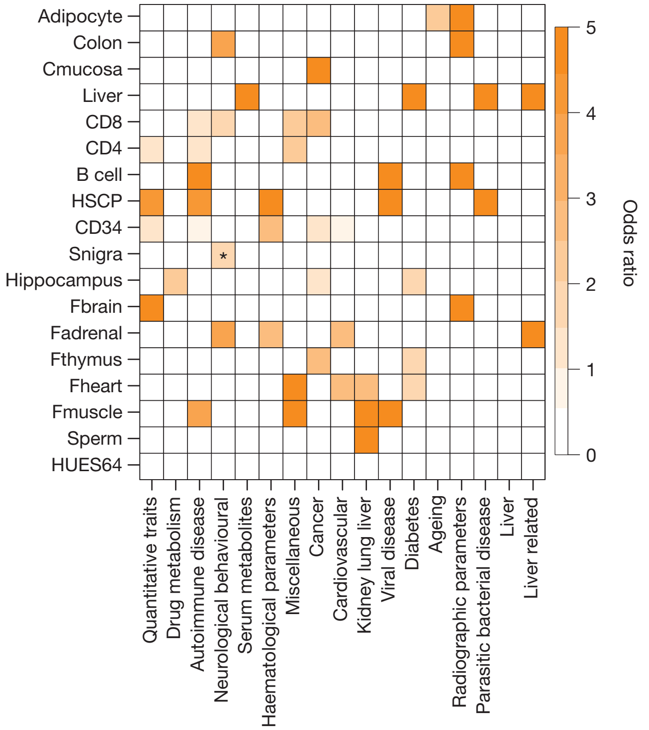


Figure 3 | DMRs exhibit increased SNP frequency and show non-random GWAS SNP enrichment. Odds ratio of significantly overrepresented ($P < 0.05$ (except for Snigra; $^*P < 0.1$), empirical test, see Supplementary Information) GWAS SNPs grouped into 16 categories in regions specifically hypomethylated within the sample indicated on the left.

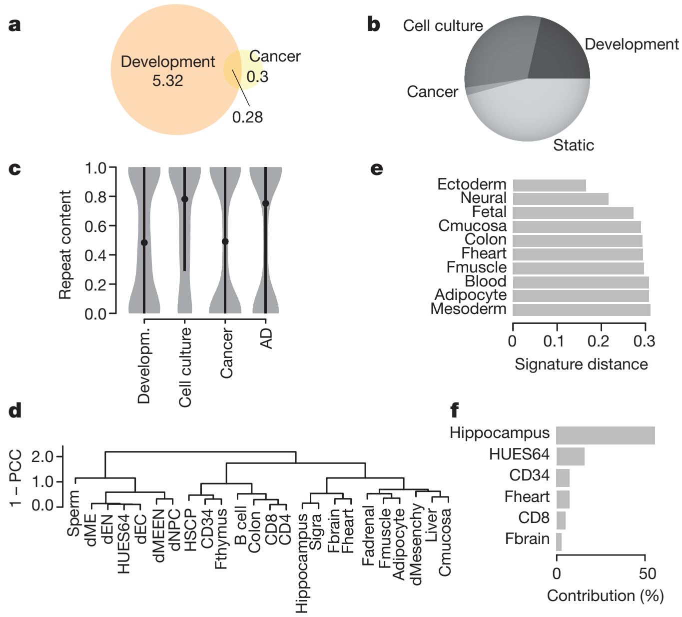


Figure 4 | Effective classification and sample deconvolution using only the DMR set. a, Overlap of dynamic CpGs ($P \leq 0.01$; $|\text{methylation difference}| \geq 0.3$) between the developmental and cancer data sets. The number of CpGs is shown in millions. **b**, Distribution of autosomal static and dynamic (across the three sets) CpGs. Class name indicates sample group in which a CpG was observed dynamic (developmental ($n = 24$), cell culture ($n = 3$), cancer ($n = 2$)) or remained unchanged over the entire sample set ($n = 30$). **c**, Repeat content distribution of DMRs (sets as in **b**). **d**, Hierarchical clustering using Pearson correlation coefficient (PCC) of the DMR values across the entire sample set ($n = 30$). **e**, Distance of the fetal brain sample to different sets of signature regions defined for sample classes or individual samples, but excluding regions identified by means of the fetal brain sample. **f**, Contribution of individual sample signature region sets to an *in-silico*-generated hybrid sample (HUES64 and hippocampus).

($n = 12{,}408$) overlapped with our previous set of developmental CpGs (4,540 out of 12,294). The most notable change in the number of dynamic CpGs occurs when comparing our developmental sample cohort to the long-term cell culture cohort, leading to the identification of 8.4 million additional dynamic CpGs (Fig. 4b). Importantly, this expanded set differs notably in terms of their sequence features, with cancer and Alzheimer's disease dynamic CpGs residing in less conserved regions that also exhibit lower motif complexity compared to the developmental and cell culture (Supplementary Fig. 4a, b). The cell-culture-specific CpGs exhibit increased repeat content relative to developmental CpGs, a feature that is shared with Alzheimer's disease (Fig. 4c). Although the disease samples clearly add more dynamic CpGs, our analysis suggests a notable overlap with our previous set for CpGs that may participate in actual regulatory events.

Finally, we investigated the utility and power of the reduced region set to classify accurately unknown samples or help to deconvolute a mixture of samples. We first clustered our developmental sample set based on the DMRs only (Fig. 4d) and found the result to be in excellent agreement with genome-wide 1-kb tiling-based clustering (Supplementary Fig. 5a). To probe the potential of our DMR set to classify unknown samples accurately, we derived signature region sets for different sample groups. These signature regions turned out to be excellent classifiers of an unseen sample (Fig. 4e, fetal brain). Next, we tested as a proof of principle whether it is possible to use our DMR set to infer the different cell populations present within a heterogeneous sample. To that end, we deconvoluted an *in silico* mixture of HUES64 and hippocampus WGBS libraries using our DNA methylation signatures. Notably, the two top hits after application of a very simple deconvolution algorithm indeed proved to be hippocampus and HUES64 (Fig. 4f).

Our study highlights and defines a relatively small subset of all genomic CpGs that change their DNA methylation state across a large number of representative cell types. Although we expect that number to increase with more diverse cell types as more WGBS data sets becoming available, our analysis suggests that the rate of newly discovered regulatory CpGs will drop rapidly once all major cell and tissue types have been mapped, mostly owed to the fact that between tissue variability exceeds within tissue variability by one order of magnitude (Supplementary Fig. 3a, b). Future studies are likely to fine map dynamics occurring in more specific subpopulations, giving rise to smaller changes in DNA methylation that we were unable to detect or include because of power constraints. Extreme conditions *in vitro* or *in vivo* such as loss or misregulation of the maintenance methylation machinery will affect a larger subset including many intergenic CpGs that are generally static, but most of these additional CpGs are unlikely to overlap with functional elements such as TFBSs or enhancers. In combination with the fact that sequencing of WGBS libraries is very inefficient, as about 65% of all 101-bp reads in our set did not even contain any CpGs to begin with, this amounts to an approximate, combined loss of around 80% of sequencing depth on non-informative reads and static regions. Furthermore, once defined, it will probably be sufficient in most cases to profile only a representative subset of CpGs across a comprehensive set of DMRs using an array-based[28] or hybrid-capture-based[29] technology to recover representative dynamics and measure regulatory events. Using these results as a guiding principle, we expect further improved efficiencies in mapping DNA methylation and enhance its applicability as a marker for various regulatory dynamics in normal and disease phenotypes.

METHODS SUMMARY

Biological materials and sequencing libraries. Genomic DNA was fragmented to 100–500 bp using a Covaris S2 sonicator. DNA fragments were cleaned-up, end-repaired, A-tailed and ligated with methylated paired-end adapters (purchased from ATDBio). See Supplementary Information for details.

Data processing and analysis. In-house-generated WGBS libraries were aligned using MAQ[30] in bisulphite mode to the hg19/GRCh37 reference assembly. Subsequently, CpG methylation calls were made using custom software, excluding duplicate,

low-quality reads, as well as reads with more than 10% mismatches. Methylation ratios of individual CpGs were described using a beta-binomial model estimating parameters from the number of methylated and total reads overlapping a given CpG, incorporating replicates. Only CpGs covered by $\geq 5\times$ reads were considered for further analysis. Differential methylation values of individual CpGs were estimated using the beta-difference distributions. CpG cluster differential methylation was determined by pooling CpG level methylation differences using a random effects model. CpG cluster methylation specificity was determined using the Jensen–Shannon divergence of the methylation level distribution of a CpG cluster across all samples, and a reference distribution representing either of the two extremes: completely unmethylated or fully methylated. *In-silico*-identified CpG islands were defined by genomic regions of at least 700 bp in length, a CpG observed versus expected ratio of greater than 0.6, and a GC content greater than or equal to 0.5. For the SNP analysis, we obtained the CEPH SNP set from the University of California Santa Cruz (USCS). GWAS SNPs were retrieved from the GWAS catalogue, whereas most of the GWAS SNP grouping was taken from ref. 24. For TFBS analysis, we retrieved peak files from the ENCODE projects and collapsed replicates. For detailed methods see Supplementary Information.

Received 15 January; accepted 4 July 2013.
Published online 7 August 2013.

1. Bestor, T. H. The DNA methyltransferases of mammals. *Hum. Mol. Genet.* **9,** 2395–2402 (2000).
2. Reik, W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447,** 425–432 (2007).
3. Seisenberger, S. *et al.* The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48,** 849–862 (2012).
4. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484,** 339–344 (2012).
5. Hackett, J. A. & Surani, M. A. DNA methylation dynamics during the mammalian life cycle. *Phil. Trans. R. Soc. B* **368,** 20110328 (2013).
6. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16,** 6–21 (2002).
7. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nature Rev. Genet.* **14,** 204–220 (2013).
8. Bergman, Y. & Cedar, H. DNA methylation dynamics in health and disease. *Nature Struct. Mol. Biol.* **20,** 274–281 (2013).
9. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).
10. Nazor, K. L. *et al.* Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell* **10,** 620–634 (2012).
11. Weber, M. *et al.* Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genet.* **37,** 853–862 (2005).
12. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454,** 766–770 (2008).
13. Laurent, L. *et al.* Dynamic changes in the human methylome during differentiation. *Genome Res.* **20,** 320–331 (2010).
14. Varley, K. E. *et al.* Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* **23,** 555–567 (2013).
15. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genet.* **41,** 178–186 (2009).
16. Berman, B. P. *et al.* Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nature Genet.* **44,** 40–46 (2012).
17. Cohen, N. M., Kenigsberg, E. & Tanay, A. Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. *Cell* **145,** 773–786 (2011).
18. Lienert, F. *et al.* Identification of genetic elements that autonomously determine DNA methylation states. *Nature Genet.* **43,** 1091–1097 (2011).
19. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489,** 75–82 (2012).
20. Zhu, J. *et al.* Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152,** 642–654 (2013).
21. Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489,** 91–100 (2012).
22. Ravasi, T. *et al.* An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* **140,** 744–752 (2010).
23. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480,** 490–495 (2011).
24. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337,** 1190–1195 (2012).
25. Bell, C. G. *et al.* Human-specific CpG "beacons" identify loci associated with human-specific traits and disease. *Epigenetics* **7,** 1188–1199 (2012).
26. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106,** 9362–9367 (2009).
27. Ehrlich, M. DNA hypomethylation in cancer cells. *Epigenomics* **1,** 239–259 (2009).
28. Dedeurwaerder, S. *et al.* Evaluation of the Infinium Methylation 450K technology. *Epigenomics* **3,** 771–784 (2011).
29. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnol.* **27,** 182–189 (2009).
30. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18,** 1851–1858 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank K. Clement, P. Samavarchi-Tehrani, Z. Smith, M. Chan and R. Karnik for discussions and feedback. We would also like to thank F. Kelley, T. Durham, C. Epstein, N. Shoresh, G. Lauwers and the Massachusetts General Hospital tissue repository for assisting in sample and data management. E.D.R. is supported by the National Institutes of Health (NIH) Roadmap Epigenomics Project (ES017690). D.A.B. is supported by NIH grants P30AG10161, R01AG17917, R01AG15819 and R01AG36042. A.M. is supported by the Pew Charitable Trusts and is a New York Stem Cell Foundation, Robertson Investigator. This work was funded by NIH grants (U01ES017155 and P01GM099117) and The New York Stem Cell Foundation.

Author Contributions M.J.Z. and A.M. conceived the study and interpreted the results. M.J.Z. designed the statistical framework, analysis strategy and analysed the data. H.G. performed in-house WGBS library production, F.M. contributed bioinformatics tools and J.D. performed cell culture experiments. L.T.-Y.T. and E.D.R. provided adipocyte nuclei for WGBS profiling, and P.L.D. and D.A.B. made adult brain and Alzheimer's disease samples available. O.K. provided support on analysis strategy and statistical methods. B.E.B. and A.M. organized samples as part of the NIH Roadmap Epigenomics Project. H.G., A.G. and A.M. established the WGBS at the Broad Institute. A.M. supervised the project. M.Z. and A.M. wrote the paper with assistance from the other authors.

Author Information WGBS data are deposited at the Gene Expression Omnibus (see Supplementary Table 1 for the specific accession numbers). Supplementary Table 2 is available under GEO accession number GSE46644. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M. (alexander_meissner@harvard.edu).

LETTER

doi:10.1038/nature12333

DNA unwinding heterogeneity by RecBCD results from static molecules able to equilibrate

Bian Liu[1,2,3], Ronald J. Baskin[2] & Stephen C. Kowalczykowski[1,2,3]

Single-molecule studies can overcome the complications of asynchrony and ensemble-averaging in bulk-phase measurements, provide mechanistic insights into molecular activities, and reveal interesting variations between individual molecules[1–3]. The application of these techniques to the RecBCD helicase of *Escherichia coli* has resolved some long-standing discrepancies, and has provided otherwise unattainable mechanistic insights into its enzymatic behaviour[4–6]. Enigmatically, the DNA unwinding rates of individual enzyme molecules are seen to vary considerably[6–8], but the origin of this heterogeneity remains unknown. Here we investigate the physical basis for this behaviour. Although any individual RecBCD molecule unwound DNA at a constant rate for an average of approximately 30,000 steps, we discover that transiently halting a single enzyme–DNA complex by depleting Mg^{2+}-ATP could change the subsequent rates of DNA unwinding by that enzyme after reintroduction to ligand. The proportion of molecules that changed rate increased exponentially with the duration of the interruption, with a half-life of approximately 1 second, suggesting that a conformational change occurred during the time that the molecule was arrested. The velocity after pausing an individual molecule was any velocity found in the starting distribution of the ensemble. We suggest that substrate binding stabilizes the enzyme in one of many equilibrium conformational sub-states that determine the rate-limiting translocation behaviour of each RecBCD molecule. Each stabilized sub-state can persist for the duration (approximately 1 minute) of processive unwinding of a DNA molecule, comprising tens of thousands of catalytic steps, each of which is much faster than the time needed for the conformational change required to alter kinetic behaviour. This ligand-dependent stabilization of rate-defining conformational sub-states results in seemingly static molecule-to-molecule variation in RecBCD helicase activity, but in fact reflects one microstate from the equilibrium ensemble that a single molecule manifests during an individual processive translocation event.

The RecBCD enzyme is an important helicase/nuclease in the repair of double-stranded DNA (dsDNA) breaks via homologous recombination[8]. RecBCD initiates homologous recombination by processing dsDNA to generate 3′-ended single-stranded DNA (ssDNA) upon recognition of the recombination hotspot sequence χ (crossover hotspot instigator (Chi); 5′-GCTGGTGG-3′[9]). The RecB and RecD subunits are SF1 helicases with 3′→5′ and 5′→3′ translocation polarities, respectively[10,11]. RecC holds the complex together and recognizes χ[12]. RecB and RecD drive dsDNA unwinding by acting as ssDNA motors, pulling the two antiparallel strands of the DNA across a pin in the RecC subunit and thus splitting the duplex DNA[13].

Earlier single-molecule studies of DNA unwinding by RecBCD revealed considerable variation in the unwinding rates of each molecule[4–7]. To understand the molecular origin of this intrinsic heterogeneity, we analysed the unwinding behaviour of a larger set of individual RecBCD molecules on bacteriophage λ DNA lacking χ (Fig. 1a, b). A total of 251 molecules were initially analysed (Fig. 1c). The majority (96%) of the molecules did not change their speeds during unwinding (Fig. 1b). Individual RecBCD molecules were observed to unwind and degrade DNA at constant velocities for 30–60 s, for over tens of thousands of catalytic turnovers. Although the rate distribution in earlier studies could be fit to a single Gaussian function[4,5], the sizes of those data sets were limited; the comparatively large number of single-molecule unwinding rates obtained here provide clear evidence of a non-unimodal distribution (Fig. 1c). The distribution was fit to the sum of two Gaussian functions; the major population of molecules (71%) has a mean fitted rate of 1,584 ± 95 base pairs (bp) s^{-1} (± standard deviation (s.d.)) whereas the minor population (29%) has a mean rate of 907 ± 500 bp s^{-1} (± s.d.). The difference in unwinding rates between the fast and slow populations is considerably beyond the experimental uncertainty. The slow population is not due to the recognition of χ-like sequences, because such events are readily discerned as pauses followed by a velocity change (Supplementary Fig. 1). Interestingly, the fast molecules are more processive than the slow ones (Supplementary Fig. 2). Both the rate and processivity of the slow species are comparable to the behaviour of RecBCD mutants with a defective motor subunit[14], leading us to examine the single-molecule behaviour of two such single-motor mutant enzymes. DNA unwinding by $RecBCD^{K177Q}$ (RecBCD* in Fig. 1c) is manifest as a single Gaussian distribution with a mean rate of 729 ± 290 bp s^{-1}, and for $RecB^{K29Q}CD$

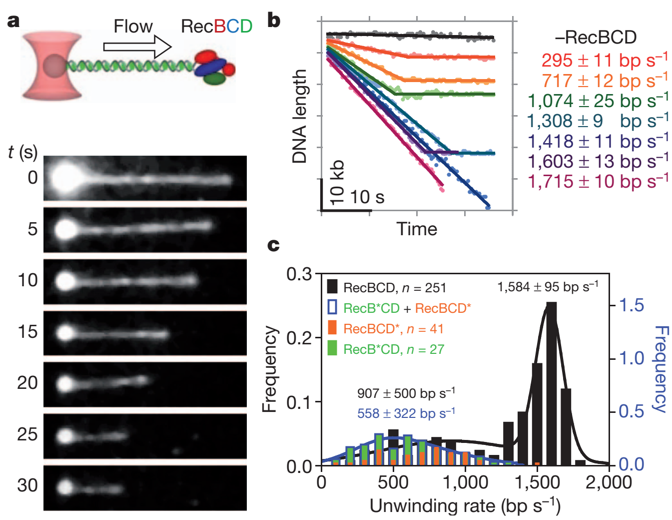


Figure 1 | Unwinding of DNA by individual RecBCD molecules is heterogeneous, with a fixed rate for the duration of DNA translocation. **a**, Visualization of a RecBCD unwinding an individual DNA molecule: experimental scheme (top) and sequential images (bottom). **b**, Time courses for unwinding DNA (lacking a χ sequence) by different RecBCD molecules: black, absence of RecBCD; colours, individual RecBCD enzymes. Errors are standard error of the fit. **c**, Distribution of unwinding rates for wild-type RecBCD and motor mutants, fit to the sum of two Gaussian functions and a single Gaussian, respectively. The distribution of the motor mutants was summed to represent equal numbers of each protein. Errors are the s.d.

[1]Department of Microbiology and Molecular Genetics, University of California, Davis, California 95616, USA. [2]Department of Molecular and Cellular Biology, University of California, Davis, California 95616, USA. [3]Biophysics Graduate Group, University of California, Davis, California 95616, USA.

(RecB*CD in Fig. 1c) it is $432 \pm 227\ \mathrm{bp\,s^{-1}}$ (see also Supplementary Fig. 2). These findings suggest that, for the wild-type enzyme, the slow species represents enzymes wherein one motor subunit is initially not engaged, but can be reversibly re-engaged when halted (see below).

The origin of heterogeneity can be dynamic or static[15–17]. Whereas dynamic heterogeneity was suggested to arise from conformational fluctuations of a protein, static heterogeneity can have different sources. It can arise from chemical heterogeneity owing to the presence of multiple related genes, or from post-translational modifications[18]. It can also result from enzyme molecules with identical chemical composition that have different stable conformational sub-states in equilibrium[15,17,19] or that are kinetically trapped in non-equilibrium states capable of multiple turnovers[20–22]. We initiated experiments designed to distinguish between these possible origins. Although the protein preparation contained no detectable heterogeneity in polypeptide composition (Supplementary Fig. 3), the distributions of unwinding rates for RecBCD eluted from different fractions of a chromatographic elution peak were examined as the first trivial source of heterogeneity; no experimentally significant differences in the distribution profiles were found (Supplementary Fig. 4). We next considered the possibility that the heterogeneity arose from RecBCD species that were not at equilibrium, but rather were trapped in different kinetic conformations. In an attempt to permit such hypothetically trapped conformations to relax to the equilibrium distribution, we subjected the enzyme population to experimental conditions that could potentially allow redistribution. Partial destabilization of protein structure, followed by refolding, can allow protein molecules to relax to their global minimum on the folding energy landscape, resulting in an equilibrium distribution of enzymes. We first used thermal annealing[23]. Ensemble assays showed that RecBCD could be heated to a maximum of 45 °C for 10 min, with no loss of activity (Supplementary Fig. 5a, b). Therefore, an enzyme population that was treated at 45 °C, and slowly cooled at a rate of 1 °C $\mathrm{min^{-1}}$, was analysed using single-molecule methods. The distribution of the rates for the thermally treated enzymes was not statistically different from the original distribution ($P = 0.45$; Supplementary Fig. 5c).

An alternative to thermal annealing is to use a chemical denaturant to unfold a protein, followed by slow removal, to permit refolding to the equilibrium distribution[24,25]. Thus, we next investigated the effect of partial unfolding of RecBCD by the classical denaturant guanidine hydrochloride (GuHCl). The enzyme could be reversibly renatured after treatment with up to 0.5 M GuHCl (Supplementary Fig. 6a). The velocity distribution of the resultant individual enzymes had a mean of $1{,}736 \pm 133\ \mathrm{bp\,s^{-1}}$ for the fast population versus $1{,}773 \pm 104\ \mathrm{bp\,s^{-1}}$ for the control (Supplementary Fig. 6b), which is the same within experimental uncertainty. The mean of the treated slow population is $556 \pm 451\ \mathrm{bp\,s^{-1}}$ versus $793 \pm 307\ \mathrm{bp\,s^{-1}}$ for the control population; although the mean for the slower group seems to be reduced, the difference is not significant ($P = 0.24$). In conclusion, neither thermal annealing nor chemical refolding produced a more homogeneous distribution, indicating that either these treatments are insufficient to permit redistribution, or that the population of RecBCD enzymes is intrinsically heterogeneous.

It remained possible that the conformational distribution of RecBCD enzyme was, in fact, at equilibrium owing to the presence of multiple conformations of similar free energy[26], but the binding of substrates could lock an enzyme in a given conformation[27,28]. For RecBCD, each DNA binding event allows unwinding of tens of thousands of base pairs, perhaps suggesting that the initial binding locks the enzyme in a conformation that lasts the duration of the unwinding process—a form of conformational selection[27]. Given that we had been unable to alter the distribution of RecBCD enzyme rates by more traditional means, we next examined whether depletion of a ligand, ATP, permitted a change to an altered conformation while bound to the DNA. Consequently, we stopped individual RecBCD molecules during the course of unwinding by depleting this essential cofactor, and then measured the rate upon reintroduction of the ligand and restarting the same enzyme. This was achieved by first moving a single, optically trapped enzyme–DNA complex into the reaction channel containing ATP to initiate unwinding. After a length of time sufficient to accurately determine the rate of DNA unwinding (~10 s), the complex was moved to a third channel that contained 10 mM EDTA, but neither Mg^{2+} nor ATP, to stop unwinding. After a defined length of time, the arrested RecBCD–DNA complex was moved back to the reaction channel to resume unwinding. By halting RecBCD in this manner for 20 s, we found that about 50% (173 out of 354) of complexes restarted unwinding when moved back to the reaction channel; we presume that RecBCD dissociated from the remainder. Fig. 2a shows the time courses for three characteristic RecBCD molecules. For molecule 1, the unwinding rate decreased from $1{,}443\ \mathrm{bp\,s^{-1}}$ to $507\ \mathrm{bp\,s^{-1}}$; for molecule 2, it was the same upon resumption; and for molecule 3, it increased from $1{,}447\ \mathrm{bp\,s^{-1}}$ to $1{,}648\ \mathrm{bp\,s^{-1}}$. After the 20-s incubation in EDTA, 53% (91 out of 173) of the molecules

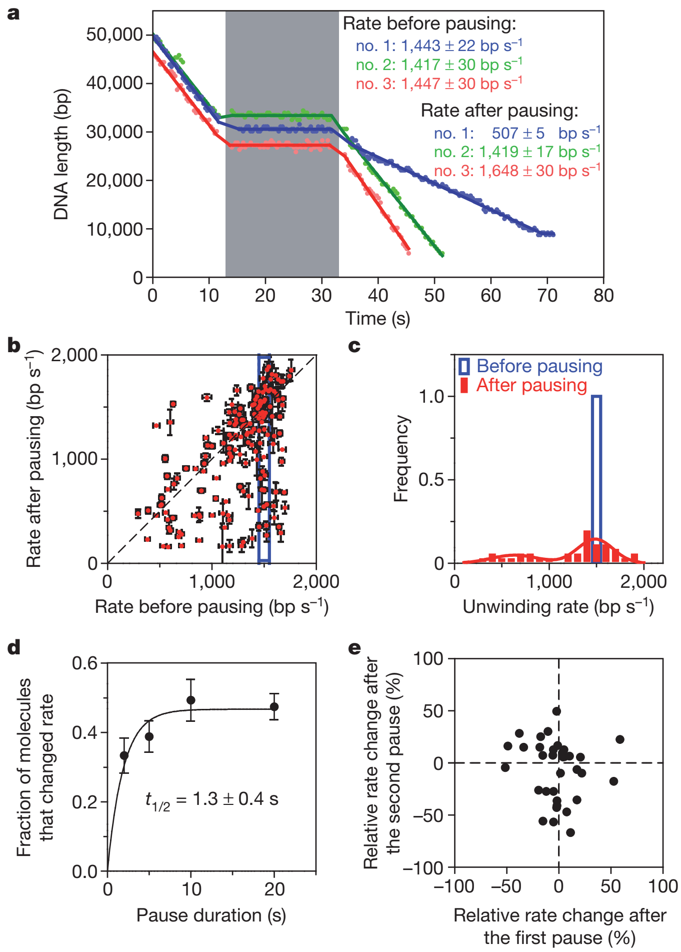


Figure 2 | The DNA unwinding rate of single enzymes is stochastically changed to a velocity within the original distribution, after transient depletion of Mg^{2+}-ATP. **a**, DNA unwinding by three representative RecBCD enzymes. The grey block indicates the pause duration. Errors are standard error of the fit. **b**, The rates before and after pausing ($n = 173$). Error bars represent the standard error of the fit. **c**, Distribution of rates before (blue) and after (red) pausing for molecules with an initial rate of $1{,}450–1{,}550\ \mathrm{bp\,s^{-1}}$ (blue box, panel **b**; $n = 36$). Before pausing, the selected bin had a mean velocity of $1{,}493 \pm 27\ \mathrm{bp\,s^{-1}}$ (s.d.); after pausing and redistribution, the mean velocity was $1{,}245 \pm 453\ \mathrm{bp\,s^{-1}}$ (s.d.) (median $= 1{,}411\ \mathrm{bp\,s^{-1}}$). **d**, Proportion of molecules that changed rates after pausing plotted versus pause duration and fitted to an exponential curve; error bars are expected bounds assuming a binomial distribution of switching events. **e**, Scatter plot of the relative rate changes after two pauses ($n = 34$).

continued unwinding with the same rate (within a 20% difference), whereas 35% ($n = 61$) of molecules slowed, and 12% ($n = 21$) of molecules increased, speed (Fig. 2b). This finding shows that the rate of individual RecBCD molecules is not static and the heterogeneity in rates cannot be, at least not solely, due to variation in covalent or irreversibly trapped structures. Note that when DNA unwinding was observed in the continuous presence of ATP (Fig. 1b), spontaneous rate-change events (Supplementary Fig. 1) were rare (4%) and attributable to χ-like recognition events. By contrast, when unwinding was interrupted by transiently removing ATP, at least 47% of the enzyme molecules resumed unwinding at a different rate upon re-introduction of ATP (Fig. 2a, b), suggesting that omission of the ATP ligand permitted a conformational switch that affects the rate-limiting translocation behaviour of RecBCD. These results support the notion that ligand binding locks the enzyme into a conformational state that typically persists for the duration of a single processive DNA unwinding transaction, whereas the absence of ATP allows the enzyme molecule to change its conformation state within the time it was halted.

The blue box in Fig. 2b highlights a binned region of the single-molecule velocity distribution containing a relatively well-populated group of molecules ($n = 36$) that translocated at rates between 1,450 and 1,550 bp s^{-1} before pausing. After incubation in EDTA, the velocities became broadly redistributed, ranging from 300 bp s^{-1} to 1,900 bp s^{-1}. The new distribution of velocities for this group is similar to the starting distribution for all the molecules (Fig. 2c and Supplementary Fig. 7), although the new distribution is overrepresented by molecules that switched to the slow macrostate (that is, with one motor disengaged). This finding demonstrates that an enzyme molecule with a fixed velocity can switch to any other velocity that was initially displayed by other enzymes in the original ensemble; similar redistribution was seen for other well-populated bins of molecules (Supplementary Fig. 7b, c). These findings indicate that all of the conformational sub-states of the ensemble are accessible to an enzyme after pausing. The velocity after ligand depletion is not related to the starting velocity of the enzyme, but rather, each enzyme equilibrated to a new velocity that was represented in the initial ensemble. The velocity distributions for enzymes, both before and after arrest, are not unimodal although, after being halted, the percentage of molecules in the slow group increases (Supplementary Fig. 7a). These results indicate that a RecBCD molecule can adopt any conformation on the free energy landscape, after being subjected to transient depletion of ATP. To ensure that the rate changes were not specific to the pausing by EDTA, experiments were conducted by stopping the RecBCD–DNA complex in a channel devoid of ATP but containing Mg^{2+}. Similar results were obtained (Supplementary Fig. 8).

When the duration of the time arrested without ATP was decreased to 2 s, the percentage of complexes that resumed unwinding increased to 78%, although fewer (33%) switched velocity (Supplementary Fig. 9). Upon increasing the incubation time in the EDTA channel, the proportion of molecules that changed rate increased exponentially with a half-life of 1.3 ± 0.4 s (Fig. 2d), suggesting that a conformational change responsible for the change in velocity in the absence of ATP requires $\sim$1 s. The combined data set for all pauses (Supplementary Fig. 9d; $n = 445$) shows that, with some underrepresentation of the slow starting velocities, there is switching from any one microstate to any other microstate. Given the existence of two macrostates (the fast population with two motors attached, and the slow population with one motor attached), when velocity switches that occur only within a macrostate are considered, the velocity redistributions are completely random (Supplementary Fig. 9e, f). Because the rate of ATP hydrolysis is rapid (ranging from a few hundred to a few thousand per s) relative to the half-life for the conformational change (1.3 s), the time between two adjacent ATP binding events would be too short (on the order of ms) for the unliganded apo-form of RecBCD to adopt a different conformation during the time that ADP has dissociated and before ATP has re-bound. For this reason, we presume that spontaneous switching is rare. Our interpretation is in accord with an earlier study which found that a few individual RecBCD enzymes can spontaneously change velocity when examined at low (15 μM) ATP[29]. At such a low ATP concentration, the apo-form of the enzyme is longer lived and the time between adjacent ATP binding events would be $\sim$67-fold longer than for the studies in this report, making it more likely that RecBCD could switch to a new conformational sub-state. Therefore, we conclude that the binding of ATP and DNA to RecBCD fixes the conformational state, which in turn defines the unwinding rate for the duration of a single processive unwinding event, contributing to the observed heterogeneity in rates of (and between) individual enzymes.

To determine whether the conformational changes are stochastic for any individual molecules, we halted some enzyme molecules twice using the same procedure, and asked whether the rate changes after each interruption were correlated. The individual molecules ($n = 34$) exhibited both decreased and increased rates after each pause, as seen above (Supplementary Fig. 10), and we found no correlation between the relative changes in rate as the results of the two consecutive pauses (Fig. 2e).

Earlier studies on the behaviour of other single enzymes have reported static heterogeneity in catalytic rates owing to variation in the covalent structures[18], the presence of metastable conformations[15,17,19] or dynamic heterogeneity caused by conformational fluctuation[16]. In this work, we found that the heterogeneity in the DNA unwinding rates by RecBCD is static on the experimental time scale of DNA unwinding for tens of thousands of base pairs. However, the rates are not intrinsic to individual molecules; thus, the heterogeneity cannot be explained by possible variations in the covalent structures of the enzyme. Instead, any individual molecule can adopt any conformation within the initially accessible free energy landscape after depletion of a ligand for a few seconds. The ergodic hypothesis posits that the (infinite) time-averaged behaviour of a molecule at equilibrium is equal to the ensemble-average of an infinite collection of those molecules. Thus, if a single enzyme molecule could be repeatedly stopped and observed, it should adopt all the possible conformations that are accessible for those conditions of thermodynamic state. Clearly, we cannot examine a single molecule for an infinite number of times, but a corollary of the ergodic hypothesis is that if one could watch any single molecule in an equilibrium distribution that could randomly switch at least once to a new conformation, then the distribution of those new states should recapitulate the original distribution, if indeed the first distribution was at equilibrium. By watching a collection of individual enzymes switching a limited number of times, here we show that they can switch to velocities found in the original distribution. Therefore, we conclude that these seemingly static RecBCD molecules can switch into microstates existing within the original ensemble. Also, when transitions remain within each macrostate, the new distribution of velocities is completely random, manifesting an expectation of ergodic behaviour. Unexpectedly, the lifetimes of these kinetic states are atypically long, and are dictated by ligand occupancy. We imagine that the conformation of the enzyme is dynamic in the absence of ligands and that a single conformation is selected and stabilized, that is, made seemingly static, upon ligand binding[27]. These findings help us to understand the influence of ligand binding on protein conformations, conformational selection and enzymatic reactions, and they now raise the intriguing structural question of how sub-states that vary in speeds by hundreds of base pairs per second can be maintained by these quasi-stable enzymatic conformations. Finally, the possible biological function of heterogeneity in a population of individual molecules is unknown and is difficult to define. However, we offer the plausible speculation that the variation seen for populations of individual molecules is akin to the epigenetic variation in the populations of organisms. Given the stochastic nature of life, a population of cells—bacteria in this specific case—needs both diversity and flexibility to respond to the random nature of natural challenges. We suggest that the variation in individual molecule behaviour affords a molecular plasticity in the cellular functions of RecBCD to respond to unpredictable needs. RecBCD has two seemingly contradictory functions: one is the degradation of

foreign duplex DNA (for example, DNA viruses) and the other is the repair of broken chromosomal DNA[8]. The regulation of these activities is controlled by recognition of the DNA regulatory sequence χ. Each *E. coli* cell contains only ten RecBCD enzyme molecules, and each cell suffers ~0.5 DNA breaks per cell cycle and is exposed to an unpredictable amount of phage or foreign DNA. If RecBCD were limited to one conformation, or if it could adopt multiple conformations but these conformations rapidly equilibrated after each step of processive unwinding, then all DNA would be processed at the same rate. Given the probabilistic nature of DNA breaks and the appearance of foreign DNA, conformational heterogeneity coupled with conformation selection of a kinetically stable functional form of RecBCD can ensure a stochastic but broad cellular response. Consequently, if the few RecBCD molecules present can adopt a wide range of conformational states, then survival through random selection is more likely, and the surviving cells, within a population of cells, are not constrained genetically. By coupling dynamic disorder in the ensemble with subsequent random selection of conformations that remain static during processive DNA unwinding, both molecules and cells can respond probabilistically to unpredictable situation with just a handful of molecules. From the perspective of a population of cells, although some will perish, a random fraction will have survived by throwing the dice productively.

METHODS SUMMARY

Single-molecule DNA helicase reactions were performed using an optical trapping and microfluidics system as reported[7,30] with minor modification. For the pausing experiments, a three-channel flow cell was used. The first channel contained bead–DNA complexes and 2 mM $Mg(OAc)_2$ in single-molecule buffer (SMB; 45 mM $NaHCO_3$ (pH 8.3), 15% (w/v) sucrose, 50 mM dithiothreitol and 20 nM YOYO-1 dye). The second channel contained 1 mM ATP and 2 mM $Mg(OAc)_2$ in SMB. The third channel contained 10 mM EDTA or 2 mM $Mg(OAc)_2$ in SMB.

For comparison of the rate distributions, the two-sample Kolmogorov–Smirnov test was used. For correlation analysis, Spearman rank correlation test was used. All P values reported for statistical analysis refer to the two-tailed probability of the tests.

Full Methods and any associated references are available in the online version of the paper.

Received 14 October 2012; accepted 22 May 2013.
Published online 14 July 2013.

1. Moffitt, J. R., Chemla, Y. R., Smith, S. B. & Bustamante, C. Recent advances in optical tweezers. *Annu. Rev. Biochem.* **77,** 205–228 (2008).
2. Ha, T. Single-molecule fluorescence resonance energy transfer. *Methods* **25,** 78–86 (2001).
3. Xie, X. S. & Lu, H. P. Single-molecule enzymology. *J. Biol. Chem.* **274,** 15967–15970 (1999).
4. Spies, M. *et al.* A molecular throttle: the recombination hotspot χ controls DNA translocation by the RecBCD helicase. *Cell* **114,** 647–654 (2003).
5. Spies, M., Amitani, I., Baskin, R. J. & Kowalczykowski, S. C. RecBCD enzyme switches lead motor subunits in response to χ recognition. *Cell* **131,** 694–705 (2007).
6. Handa, N., Bianco, P. R., Baskin, R. J. & Kowalczykowski, S. C. Direct visualization of RecBCD movement reveals cotranslocation of the RecD motor after χ recognition. *Mol. Cell* **17,** 745–750 (2005).
7. Bianco, P. R. *et al.* Processive translocation and DNA unwinding by individual RecBCD enzyme molecules. *Nature* **409,** 374–378 (2001).
8. Dillingham, M. S. & Kowalczykowski, S. C. RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol. Mol. Biol. Rev.* **72,** 642–671 (2008).
9. Lam, S. T., Stahl, M. M., McMilin, K. D. & Stahl, F. W. Rec-mediated recombinational hot spot activity in bacteriophage lambda. II. A mutation which causes hot spot activity. *Genetics* **77,** 425–433 (1974).
10. Dillingham, M. S., Spies, M. & Kowalczykowski, S. C. RecBCD enzyme is a bipolar DNA helicase. *Nature* **423,** 893–897 (2003).
11. Taylor, A. F. & Smith, G. R. RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity. *Nature* **423,** 889–893 (2003).
12. Handa, N. *et al.* Molecular determinants responsible for recognition of the single-stranded DNA regulatory sequence, χ, by RecBCD enzyme. *Proc. Natl Acad. Sci. USA* **109,** 8901–8906 (2012).
13. Singleton, M. R., Dillingham, M. S., Gaudier, M., Kowalczykowski, S. C. & Wigley, D. B. Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature* **432,** 187–193 (2004).
14. Dillingham, M. S., Webb, M. R. & Kowalczykowski, S. C. Bipolar DNA translocation contributes to highly processive DNA unwinding by RecBCD enzyme. *J. Biol. Chem.* **280,** 37069–37077 (2005).
15. Frauenfelder, H., McMahon, B. H., Austin, R. H., Chu, K. & Groves, J. T. The role of structure, energy landscape, dynamics, and allostery in the enzymatic function of myoglobin. *Proc. Natl Acad. Sci. USA* **98,** 2370–2374 (2001).
16. Lu, H. P., Xun, L. & Xie, X. S. Single-molecule enzymatic dynamics. *Science* **282,** 1877–1882 (1998).
17. Xue, Q. & Yeung, E. S. Differences in the chemical reactivity of individual molecules of an enzyme. *Nature* **373,** 681–683 (1995).
18. Craig, D. B., Arriaga, E. A., Wong, J. C. Y., Lu, H. & Dovichi, N. J. Studies on single alkaline phosphatase molecules: reaction rate and activation energy of a reaction catalyzed by a single molecule and the effect of thermal denaturation – the death of an enzyme. *J. Am. Chem. Soc.* **118,** 5245–5253 (1996).
19. Shi, J. *et al.* Multiple states of the Tyr318Leu mutant of dihydroorotate dehydrogenase revealed by single-molecule kinetics. *J. Am. Chem. Soc.* **126,** 6914–6922 (2004).
20. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. Navigating the folding routes. *Science* **267,** 1619–1620 (1995).
21. Onuchic, J. N., Wolynes, P. G., Luthey-Schulten, Z. & Socci, N. D. Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl Acad. Sci. USA* **92,** 3626–3630 (1995).
22. Dill, K. A., Ozkan, S. B., Shell, M. S. & Weikl, T. R. The protein folding problem. *Annu. Rev. Biophys.* **37,** 289–316 (2008).
23. Nguyen, H. D. & Hall, C. K. Effect of rate of chemical or thermal renaturation on refolding and aggregation of a simple lattice protein. *Biotechnol. Bioeng.* **80,** 823–834 (2002).
24. Ikai, A. & Tanford, C. Kinetic evidence for incorrectly folded intermediate states in the refolding of denatured proteins. *Nature* **230,** 100–102 (1971).
25. Sela, M., White, F. H. Jr & Anfinsen, C. B. Reductive cleavage of disulfide bridges in ribonuclease. *Science* **125,** 691–692 (1957).
26. Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254,** 1598–1603 (1991).
27. Ma, B. & Nussinov, R. Enzyme dynamics point to stepwise conformational selection in catalysis. *Curr. Opin. Chem. Biol.* **14,** 652–659 (2010).
28. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chem. Biol.* **5,** 789–796 (2009).
29. Perkins, T. T., Li, H. W., Dalal, R. V., Gelles, J. & Block, S. M. Forward and reverse motion of single RecBCD molecules on DNA. *Biophys. J.* **86,** 1640–1648 (2004).
30. Amitani, I., Liu, B., Dombrowski, C. C., Baskin, R. J. & Kowalczykowski, S. C. Watching individual proteins acting on single molecules of DNA. *Methods Enzymol.* **472,** 261–291 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to members of the laboratory for their comments on this work. S.C.K. was supported by the National Institutes of Health (GM-62653 and GM-64745).

Author Contributions B.L., R.J.B. and S.C.K. conceived the general ideas, designed the experiments and interpreted the data. B.L. performed experiments. B.L. and S.C.K. analysed the data and wrote the manuscript. R.J.B. passed away on July 3, 2010; this work is dedicated to his collegiality and contributions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.C.K. (sckowalczykowski@ucdavis.edu).

METHODS

Proteins and DNA substrates. *E. coli* RecBCD enzyme was expressed and purified as described previously[31,32]. To check purity, protein was analysed using a 12% denaturing polyacrylamide gel (1:29 bis:acrylamide in TBE buffer (89 mM Tris base, 2 mM EDTA, 89 mM boric acid), containing 10% SDS) stained with Coomassie blue dye. After electrophoresis, the gel was imaged using an AlphaInnotech gel documentation system. The two mutant enzymes, $RecBCD^{K177Q}$ (RecBCD*) and $RecB^{K29Q}CD$ (RecB*CD), were purified as described[10].

Bacteriophage λ DNA (N^6-methyladenine-free lambda DNA, New England Biolabs) was biotinylated by ligation to a 3′-biotinylated 12-mer oligonucleotide (*cosA*: 5′-GGGCGGCGACCT-3′ or *cosB*: 5′-AGGTCGCCGCCC-3′, Operon Technologies) that is complementary to one of the cohesive ends of λ DNA[7]; except for the thermal re-annealing and control experiments, where *cosA* was used, all other experiments used the *cosB* oligonucleotide.

The pUC19 plasmid DNA was purified by caesium chloride gradient centrifugation. The circular DNA was linearized with NdeI restriction endonuclease (New England Biolabs) followed by heat inactivation and phenol/chloroform/isoamyl alcohol extraction. The DNA concentration was determined by absorbance at 260 nm using an extinction coefficient of 6,330 M^{-1} (nucleotides) cm^{-1}.

ATP hydrolysis assays. The ATP hydrolysis activity of the enzyme was measured spectrophotometrically as reported[33] by coupling ATP hydrolysis to NADH oxidation[34] using an Agilent Technologies Model 8452A diode array spectrophotometer. The assay mixtures contained 25 mM Tris acetate (pH 7.5), 1 mM dithiothreitol (DTT), 2 mM ATP, 5 mM magnesium acetate, 1.5 mM phosphoenolpyruvate, 0.2 mg ml^{-1} NADH, 30 U ml^{-1} pyruvate kinase, 30 U ml^{-1} lactate dehydrogenase and 50 μM (nucleotides) poly(dT). Reactions were initiated by the addition of 0.5 nM RecBCD enzyme after pre-incubation of all other components at 37 °C for 5 min. The rate of ATP hydrolysis was calculated from the rate of change in absorbance at 340 nm due to oxidation of NADH using the following conversion: rate of $A_{340\,nm}$ decrease $(s^{-1}) \times 9{,}820 \div 0.0005$ (μM RecBCD) $\div 60 =$ rate of ATP hydrolysis (s^{-1})[33].

Re-purification of RecBCD. RecBCD enzyme (0.1 mg) from −80 °C freezer stock was thawed on ice, diluted fivefold using cold B100 buffer (20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 0.1 mM DTT and 100 mM NaCl), and loaded onto a 1-ml MonoQ column (Amersham Biosciences). The enzyme was eluted using a gradient from 300 mM to 450 mM NaCl in 30 column volumes. Three fractions (100 μl each) on one side of the peak in ultraviolet absorbance were used immediately for single-molecule helicase assays.

Stopped-flow dye-displacement helicase assay. Essentially, the protocols used previously[14] were followed. Experiments were performed in an Applied Photophysics SX.18MV-R stopped-flow apparatus with excitation at 355 nm (bandwidth 9.3 nm) and emission was measured using a 450 nm long-pass filter. Unless stated otherwise, all reported concentrations are final after mixing of equal volumes in the stopped-flow apparatus. Reactions were performed at 25 °C in a buffer containing 25 mM Tris acetate (pH 7.5), 6 mM magnesium acetate, 1 mM DTT, 200 nM Hoechst 33258 dye (Molecular Probes) and 300 nM ssDNA-binding protein (SSB). The RecBCD enzyme, at the final concentration indicated, was incubated with 0.05 nM (molecules) NdeI-cut pUC19 DNA (equivalent to 0.1 nM RecBCD binding sites) for 5 min, and this was then mixed with 2 mM ATP to initiate the reaction. Data were analysed using GraphPad Prism 5.02 (GraphPad Software). Unwinding rates were determined by a linear fit to the first 2 s of each trace.

Thermal treatment of RecBCD. Aliquots of the RecBCD enzyme in storage buffer (20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 0.1 mM DTT, 100 mM NaCl and 50% (v/v) glycerol) were thawed on ice and then heated to 45 °C for 10 min followed by slowly cooling by 1 °C min^{-1} down to 4 °C using GeneMate PCR machine. The untreated controls were kept on ice until use.

Chemical unfolding of RecBCD. Aliquots of the RecBCD enzyme were thawed on ice. Various concentrations of GuHCl were mixed in 1:1 volume ratio with the enzyme. After incubation at room temperature (∼23 °C) for 1 h, the sample was dialysed against B100 buffer (20 mM Tris-HCl (pH 7.5), 0.1 mM EDTA, 0.1 mM DTT and 100 mM NaCl) at 4 °C for 24 h and the dialysis buffer was changed once. The next day, samples were collected and the concentrations were measured after centrifugation. Samples were taken for ATPase assays, and the rest were used for single-molecule assays.

Optical trapping and fluorescence microscopy. Single-molecule DNA helicase reactions were performed using an optical trapping system as reported[7] with some modifications[30]. The system is constructed around a Nikon Eclipse microscope (Nikon). A high-pressure mercury lamp (100 W; USHIO America) and Y-FL 4-cube Epi-Fluorescence (Nikon) attachment were used for illumination. Images were captured using a high sensitivity electron bombardment couple-charged device (CCD) camera (EB-CCD C7190; Hamamatsu Photonics) and digitalized online using an LG-3 frame grabber (Scion Corporation) at 30 frames s^{-1}. The optical trap was created by focusing a 1,064 nm laser (Nd:YVO_4, 6 W max, J-series power supply; Spectra Physics) through a high numerical aperture (NA) objective (×100/1.3 oil DICH; Nikon). A high NA objective is necessary to create an intensity gradient sufficiently large to form the trap[35]. The laser is expanded with a 20× beam expander (HB-20XAR.33; Newport) to fill the back aperture of the objective. The laser is collimated and aligned using a 1× telescope. The laser is reflected along the optical axis of the microscope by means of a low-pass dichroic mirror placed between the objective and the fluorescence cube.

Experiments were carried out in a multi-channel microfluidic flow cell secured on a computer controlled motorized stage (MS-2000; Applied Scientific Instruments) mounted on the microscope. The design of the flow cell allows laminar flow of different solutions without mixing. The solutions are introduced into the flow cell by a syringe pump with multiple syringes (KD Scientific), generating a flow rate of ∼100–150 μm s^{-1}. PEEK tubing (Upchurch Scientific) is used to connect the syringes to the flow cell. The microfluidic system permits imaging of protein–DNA complexes on a single molecule of flow stretched DNA; it also enables the rapid movement of the sample to the different buffers in the channels of the flow cell. The position of the stage and, hence the flow cell, is controlled using a custom-built program. Bead–DNA complexes can be moved between adjacent solution channels within 1 s via the movement of the stage. For the pausing experiments, a three-channel flow cell was used.

Single-molecule DNA helicase reactions. The protocol used for DNA–bead preparation was modified from that used previously[4,7]. Biotinylated λ DNA (100 pM in 1–2 μl) was incubated with 1–2 μl of 1 μm ProActive streptavidin-coated microspheres (∼35 pM; Bangs Laboratories) for 1 h on ice or at 37 °C. Bead–DNA complexes were then transferred into 0.5 ml of de-gassed sample solution containing 45 mM $NaHCO_3$ (pH 8.3), 20% (w/v) sucrose, 50 mM DTT and 20 nM YOYO-1 dye (Molecular Probes). DNA was incubated with the dye for at least 1 h in the dark at room temperature. Immediately before transfer to the sample syringe, magnesium acetate and RecBCD, to final concentrations of 2 mM and 10–60 nM, respectively, were added to the sample mixture. For the control, the RecBCD storage buffer without RecBCD was used to replace the enzyme solution. The reaction solution contained 45 mM $NaHCO_3$ (pH 8.3), 20% (w/v) sucrose, 50 mM DTT, 1 mM ATP, 2 mM magnesium acetate and 20 nM YOYO-1 dye. For the pausing experiments, the three-channels were as follows. The first channel contained bead–DNA complexes and 2 mM $Mg(OAc)_2$ in SMB (45 mM $NaHCO_3$ (pH 8.3), 15% (w/v) sucrose, 50 mM dithiothreitol and 20 nM YOYO-1 dye). The second channel contained 1 mM ATP and 2 mM $Mg(OAc)_2$ in SMB. The third channel contained either 10 mM EDTA or 2 mM $Mg(OAc)_2$ in SMB; the two solutions used as indicated were either: 45 mM $NaHCO_3$ (pH 8.3), 15% (w/v) sucrose, 50 mM DTT, 10 mM EDTA and 20 nM YOYO-1 dye, or 45 mM $NaHCO_3$ (pH 8.3), 15% (w/v) sucrose, 50 mM DTT, 2 mM magnesium acetate and 20 nM YOYO-1 dye.

Single-molecule data analysis. Videos were digitalized through an LG-3 frame-grabber card using an ImageJ plugin. Images were then averaged and the length of the DNA molecule in each frame was measured using a custom-built ImageJ plugin[30]. The experimental data were fitted to either a line or a three-segment line using Origin 7.5 (OriginLab Corp.) or GraphPad Prism 5.02 (GraphPad Software, Inc.). The translocation rates of RecBCD were calculated from the slopes of the corresponding segments and the standard error of the best-fit values are reported. Unless otherwise indicated, standard deviation is reported for statistical analysis of a number of molecules. The analysis method has an estimated resolution of 50 bp s^{-1}. The difference in the unwinding rates between the fast and slow populations is significantly beyond the experimental uncertainty. When a distribution of unwinding rates was plotted, the rates were grouped in 100 bp s^{-1} bins. The distributions were fit to the sum of two Gaussian curves, unless otherwise noted. Error bars in Fig. 2d represent the expected bounds assuming a binomial distribution of switching events for the given sample size. For comparison of the rate distributions, the two-sample Kolmogorov–Smirnov test was used. For correlation analysis, Spearman rank correlation test was used. All *P* values reported for statistical analysis refer to the two-tailed probability of the tests.

31. Roman, L. J. & Kowalczykowski, S. C. Characterization of the helicase activity of the *Escherichia coli* RecBCD enzyme using a novel helicase assay. *Biochemistry* **28,** 2863–2873 (1989).
32. Bianco, P. R. & Kowalczykowski, S. C. The recombination hotspot Chi is recognized by the translocating RecBCD enzyme as the single strand of DNA containing the sequence 5′-GCTGGTGG-3′. *Proc. Natl Acad. Sci. USA* **94,** 6706–6711 (1997).
33. Spies, M., Dillingham, M. S. & Kowalczykowski, S. C. Translocation by the RecB motor is an absolute requirement for χ−recognition and RecA protein loading by RecBCD enzyme. *J. Biol. Chem.* **280,** 37078–37087 (2005).
34. Kreuzer, K. N. & Jongeneel, C. V. *Escherichia coli* phage T4 topoisomerase. *Methods Enzymol.* **100,** 144–160 (1983).
35. Neuman, K. C. & Block, S. M. Optical trapping. *Rev. Sci. Instrum.* **75,** 2787–2809 (2004).

LETTER

doi:10.1038/nature12327

Structural basis for molecular recognition of folic acid by folate receptors

Chen Chen[1,2]*, Jiyuan Ke[1]*, X. Edward Zhou[1], Wei Yi[3], Joseph S. Brunzelle[4], Jun Li[5], Eu-Leong Yong[5], H. Eric Xu[1,3] & Karsten Melcher[1]

Folate receptors (FRα, FRβ and FRγ) are cysteine-rich cell-surface glycoproteins that bind folate with high affinity to mediate cellular uptake of folate. Although expressed at very low levels in most tissues, folate receptors, especially FRα, are expressed at high levels in numerous cancers to meet the folate demand of rapidly dividing cells under low folate conditions[1–3]. The folate dependency of many tumours has been therapeutically and diagnostically exploited by administration of anti-FRα antibodies, high-affinity antifolates[4,5], folate-based imaging agents and folate-conjugated drugs and toxins[6–8]. To understand how folate binds its receptors, we determined the crystal structure of human FRα in complex with folic acid at 2.8 Å resolution. FRα has a globular structure stabilized by eight disulphide bonds and contains a deep open folate-binding pocket comprised of residues that are conserved in all receptor subtypes. The folate pteroate moiety is buried inside the receptor, whereas its glutamate moiety is solvent-exposed and sticks out of the pocket entrance, allowing it to be conjugated to drugs without adversely affecting FRα binding. The extensive interactions between the receptor and ligand readily explain the high folate-binding affinity of folate receptors and provide a template for designing more specific drugs targeting the folate receptor system.

Folates (vitamin B_9) are important one-carbon donors for the synthesis of purines and thymidine—essential components of nucleic acids—and indirectly, via *S*-adenosyl methionine, for methylation of DNA, proteins and lipids[9]. Folate deficiency is therefore associated with many diseases, including fetal neural tube defects, cardiovascular disease and cancers[10]. In adult tissues, folate is mainly taken up by reduced folate carrier, a ubiquitously expressed anion channel that has relatively low folate-binding affinity ($K_m = 1$–$10\,\mu$M)[11]. By contrast, high-affinity uptake of the food supplement folic acid ($K_d < 1$ nM)[12] and the physiologically prevalent folate N^5-methyltetrahydrofolate (5-mTHF) requires the function of three subtypes of folate receptor (FRα, FRβ and FRγ), which are cysteine-rich glycoproteins that mediate folate uptake through endocytosis. Inside of the cell, the acidic environment of the endosome promotes the release of folate from receptors, which is then transported into the cytoplasm by proton-coupled folate transporter[13]. The expression of folate receptors is largely restricted to cells important for embryonic development (for example, placenta and neural tubes) and folate resorption (kidney). Among the three FR isoforms, FRα is the most widely expressed, with very low levels in normal tissues, but high expression levels in many tumours[14]. As

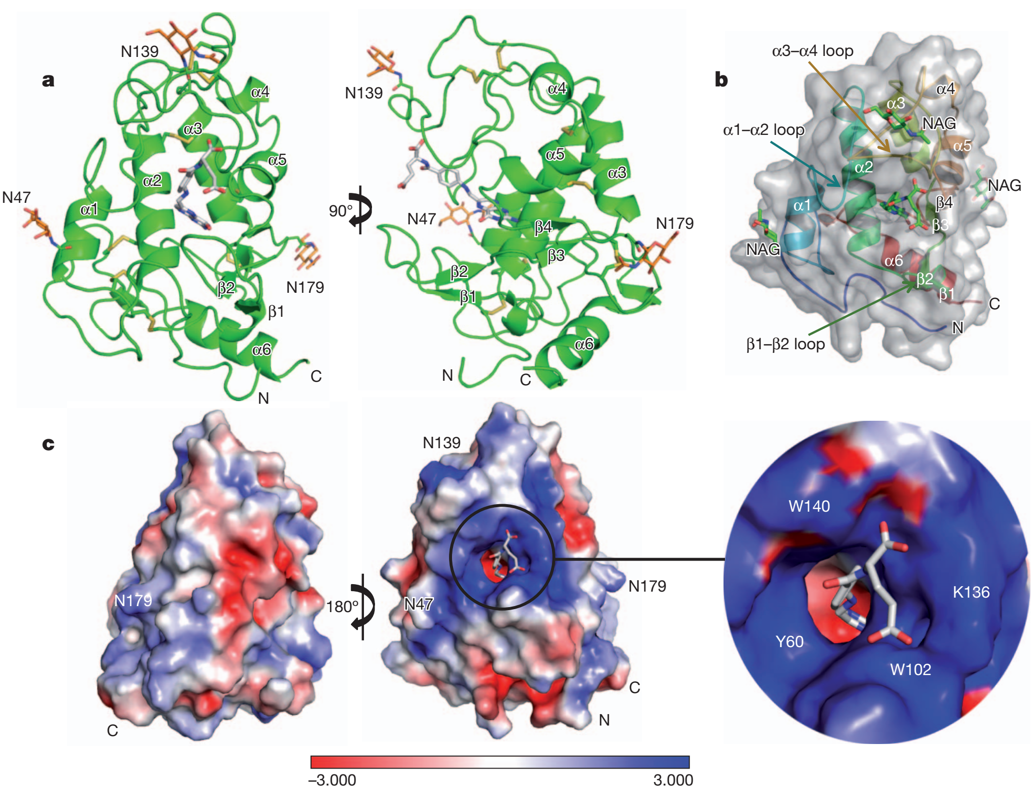


Figure 1 | Structure of FRα bound to folic acid. a, Two views of the complex, with FRα in green, folic acid in grey, NAG in orange and the disulphide bonds depicted as yellow sticks. The N and C termini are labelled. **b**, Ribbon diagram of FRα, with folic acid and NAG in green stick presentations, overlaid with the semi-transparent receptor surface. **c**, Charge distribution surface of FRα with a close-up view of the ligand-binding pocket entrance. Folic acid carbon atoms are coloured grey, nitrogen atoms blue, and oxygen atoms red. A colour-code bar (bottom) shows an electrostatic scale from −3 to +3 eV.

[1]Program for Structural Biology and Drug Discovery, Van Andel Research Institute, 333 Bostwick Avenue North East, Grand Rapids, Michigan 49503, USA. [2]National University of Singapore Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore 117456, Singapore. [3]VARI/SIMM Center, Center for Structure and Function of Drug Targets, CAS-Key Laboratory of Receptor Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China. [4]Life Sciences Collaborative Access Team, Synchrotron Research Center, Northwestern University, Argonne, Illinois 60439, USA. [5]Department of Obstetrics & Gynecology, National University Hospital, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119074, Singapore.
*These authors contributed equally to this work.

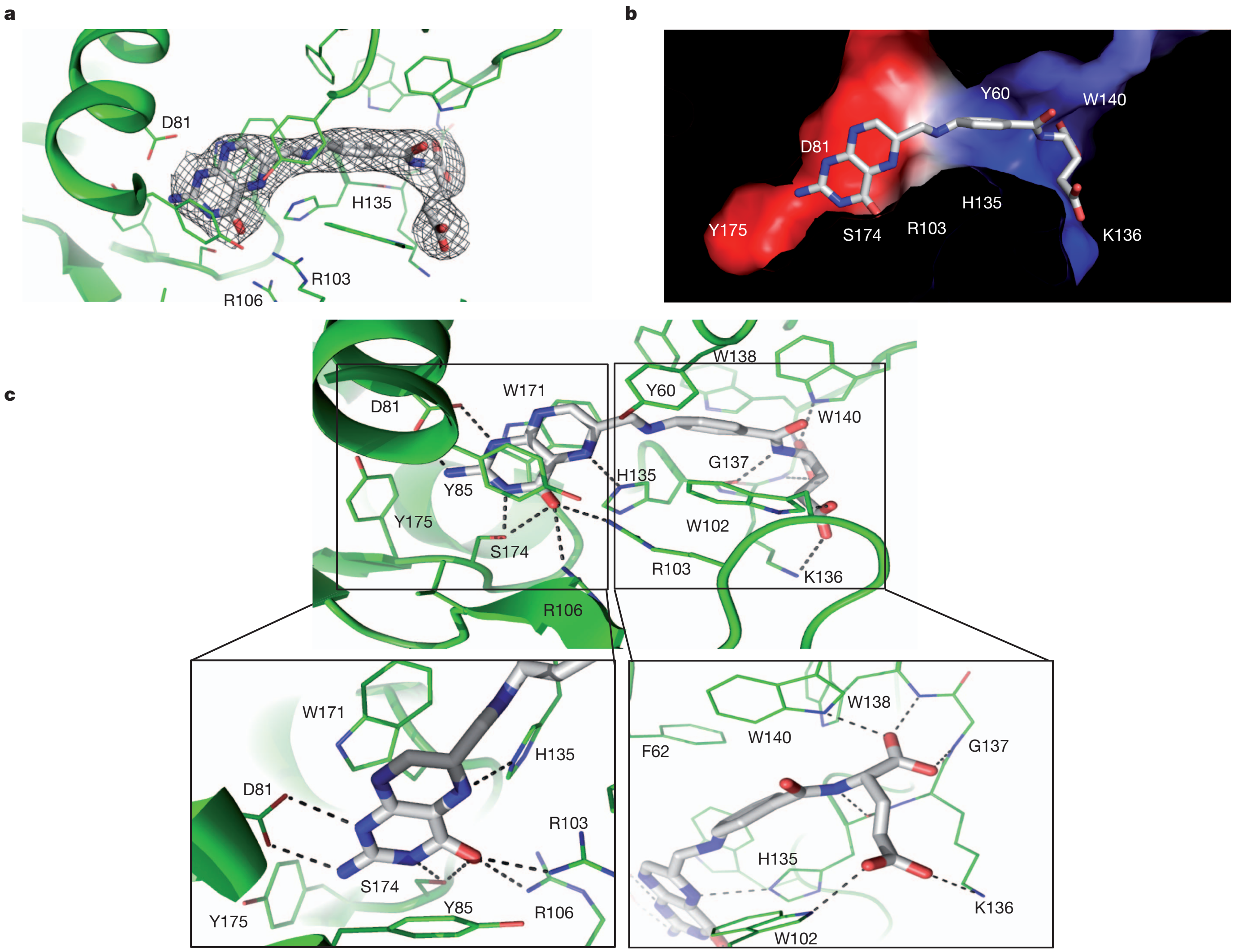


Figure 2 | Structural and biochemical analysis of FRα–folic acid interactions. **a**, The σA-weighted $2F_o - F_c$ electron density map for folic acid, shown as a grey mesh. **b**, The internal charge distribution surface of the binding pocket is shown using the same colour code as in Fig. 1c, with folic acid shown in stick presentation. **c**, Folic acid-binding network with close-ups of the folic acid head and tail groups. Residues that line the binding pocket are shown in green and folic acid is shown in grey. Hydrogen bonds are indicated by dashed lines.

such, FRα has become the molecular target for the development of many cancer therapeutics, including anti-FRα antibodies, high-affinity anti-folates, folate-based imaging agents and folate-conjugated drugs and toxins. Despite intense research on the folate structure–activity relationship, the molecular basis for the high-affinity recognition of folates by FRα remains elusive owing to the technical difficulties in expression, purification and crystallization of FRα for structural studies.

To obtain FRα protein for structural studies, we stably expressed human FRα lacking its carboxy-terminal glycophosphatidylinositol anchor as a secreted IgG Fc fusion protein (FRα–Fc) in HEK293 cells. As fully glycosylated fusion protein purified from culture medium yielded poorly diffracting crystals, we reduced crystallization-inhibiting glycosylation heterogeneity by combined treatment with kifunensine and endoglycosidase H, which together reduce complex carbohydrates to single *N*-acetylglucosamine (NAG) moieties[15] (Supplementary Fig. 1a, b). The deglycosylated FRα–Fc had a similar folic acid-binding affinity (~190 pM) to the fully glycosylated protein (Supplementary Fig. 1d) and yielded crystals, which diffracted to 2.8 Å (Supplementary Fig. 1c). We solved the structure by combining the phase information from one Pt derivative and six native S anomalous data sets (see Methods and Supplementary Table 1).

FRα has an overall globular structure, comprising four long α-helices (α1, α2, α3, α6), two short α-helices (α4, α5), four short β-strands (β1–β4) and many loop regions (Fig. 1a, b). The tertiary structure is greatly stabilized by eight disulphide bonds formed by 16 conserved cysteine residues (C15–C43, C35–C83, C44–C87, C67–C153, C74–C124, C113–C187, C117–C167 and C130–C147). FRα has three predicted *N*-glycosylation sites at N47, N139 and N179. Clear electron density for NAG is observed for N47 and N139, and partial electron density for N179. The overall fold of FRα is similar to that of riboflavin-binding protein (22% sequence identity to FRα)[16], with a root mean squared deviation of 1.56 Å for 163 Cα atoms, but the two proteins have very differently shaped ligand pockets and ligand-binding modes (Supplementary Fig. 2).

The core domain consists of helices α1, α2, α3 and α5 tied together by four disulphide bridges (C35–C83, C44–C87, C74–C124 and C117–C167; Fig. 1a). The structure of FRα contains a long and open folate-binding pocket, which is formed by α1, α2 and α3 in the back; the amino-terminal loop, β1 and β2 in the bottom; the α1–α2 and α3–α4 loops in the left and top; and α4, α5, β4 and β3 in the right (Fig. 1a, b). Folic acid is oriented with its basic pteroate moiety docked deep inside of the negatively charged pocket and the two negatively charged carboxyl groups of its glutamate moiety sticking out of the positively charged

entrance of the ligand-binding pocket, which is formed by the α1–α2, β1–β2 and α3–α4 loops (Figs 1 and 2b).

Clear electron density was observed for folic acid and the surrounding amino acid residues, which allowed for accurate modelling of the ligand and its interacting residues lining the binding pocket (Fig. 2a and Supplementary Fig. 3). The cross-section of the binding pocket reveals the complementary shape and charge between the bound ligand and the receptor (Fig. 2b). Folic acid docks into an extended groove of FRα in the direction roughly perpendicular to the plane formed by helices α1, α2 and α3, with the pterin head group buried inside against the back formed by α1, α2 and α3 (Figs 1a and 2b, c). The interactions around the pteroate moiety contain both hydrogen bonds and hydrophobic interactions. First, the pterin ring is stacked between the parallel side chains of Y85 and W171, and capped by Y175 (Fig. 2c). Second, the hydrophilic pterin ring N and O atoms form a series of hydrogen bonds with receptor residues. Specifically, the pterin N1 and N2 atoms form strong hydrogen bonds with the side-chain carboxyl group of D81, the N3 and O4 atoms with the S174 hydroxyl group, the O4 atom forms two hydrogen bonds with the guanidinium groups of R103 and R106, and the N5 atom forms one hydrogen bond with the H135 side chain (Figs 2c and 3a). Interestingly, folic acid O4 is replaced by an amino group in the antifolates methotrexate and aminopterin, which have reduced affinity for FRα[5,17]. The amino group would not allow for the formation of hydrogen bonds with R103 and R106 and would sterically clash with the position of R103 (see Fig. 3a) in the folic acid-bound structure, providing a structural rationale for the poor FRα-binding of these two compounds and their preferential uptake by reduced folate carrier.

The folic acid aminobenzoate is stabilized by hydrophobic interactions with Y60, W102 and W134, which line the middle of the long ligand-binding pocket (Fig. 3a). Extensive interactions are also observed for the glutamate group, which engages six hydrogen bonds, contributed by the side chains of W102, K136 and W140, as well as by backbone interactions with H135, G137 and W138 (Figs 2c and 3a). Most residues involved in ligand binding are identical among different subtypes of FR regardless of their origins (Supplementary Figs 4 and 5), indicating that the observed folate-binding interactions are probably conserved in all three different receptor subtypes. In addition, the most physiologically prevalent folate, 5-mTHF, can be easily docked into the FRα ligand-binding pocket in a mode very similar to that of folic acid, suggesting that the fundamental mechanism of folate recognition is conserved (Supplementary Fig. 6).

To validate the structure observations, we examined the ligand-binding affinities of FRα mutants that have alanine mutations in the key folate-contacting residues. The W171A mutation abolished the expression of the receptor (Supplementary Fig. 7a), suggesting that this residue is critical for protein stability. All other mutants expressed relatively well and were purified to determine their folate-binding affinity by radioligand-binding assay (Supplementary Figs 7b and 8b). Whereas wild-type FRα bound to [^{3}H]-folic acid with a K_d of ~0.19 nM, replacement of D81 decreased affinity by more than one order of magnitude, consistent with the strong interaction of the aspartate carboxyl oxygens with the pterin N1 and N2 nitrogens, and indicating that this interaction is a key contributor to high-affinity ligand binding. By contrast, mutations of Y175, K136 and R106 (bond lengths ≥ 3.1 Å) have little effect, and mutations of all other ligand-binding residues (hydrogen bonds ≤ 3.0 Å) have only moderate effects on folic acid binding (affinity deceases of ≤ 3.6-fold), which are approximately additive for the double mutants R103A/S174A and W102A/R103A. This extensive interaction network therefore makes FRα–folic acid binding remarkably resistant to single amino acid substitutions (Fig. 3b and Supplementary Figs 7b and 8b). Together, the structural and mutational analyses present a structural rationale for the absolute requirement of the pterin group for anchoring folate in the binding pocket of the receptor and for the availability of the glutamate group

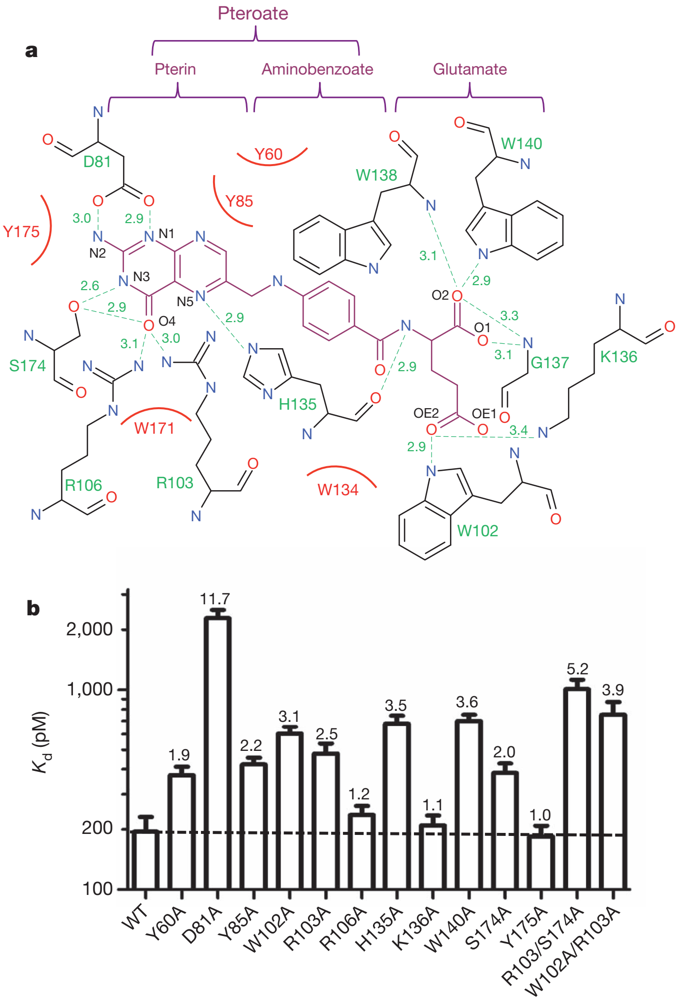


Figure 3 | Folic acid affinities of FRα ligand-binding-pocket mutants. **a**, Interaction map of folic acid with ligand-binding-pocket residues. The folic acid chemical structure is shown in magenta, pocket residues in black and hydrogen bonds as green dashed lines with bond distances (Å) indicated. Hydrophobic interactions are presented as curved red lines. The pteroate and glutamate moieties of folic acid are indicated above the map. **b**, Folic acid affinities of wild-type and mutant FRα proteins as measured by [^{3}H]-folic acid binding assay (see Supplementary Figs 7 and 8 for binding isotherms). The numbers on top of the bars indicate the fold decrease in affinity (increase in K_d) relative to wild-type FRα. Error bars indicate s.d. ($n = 2$).

for conjugation with drugs and imaging reagents[18], without adversely affecting the interactions between receptor and ligand.

In summary, many cancers highly express FRα, which has therefore become an important target for receptor-mediated chemotherapy. How FR binds to folate and folate-conjugated drugs, however, has remained unknown. The FRα–folic acid complex structure illustrates how the receptor assumes a deep folate-binding pocket that is formed by conserved residues across all receptor subtypes and provides detailed insights into how folic acid interacts with its receptors. Together, these observations establish a rational foundation for designing specific drugs targeting the folate receptor system.

METHODS SUMMARY

FRα was stably expressed in HEK293 cells as a secreted IgG Fc fusion protein in cell medium supplemented with folic acid and kifunensine. Proteins were purified from the conditioned cell media by nickel-nitrilotriacetic acid chromatography (Ni-NTA), followed by proteolytic release of the Fc tag, enzymatic deglycosylation and size-exclusion chromatography. Crystals were grown by vapour diffusion. One native data set, one Pt derivative data set and six S anomalous signal data sets were collected from cryo-protected crystals at beamlines 21-ID-D at the Advanced Photon Source at Argonne National Laboratories.

Initial phases were established using the SHELX program[19] with Pt-soaked derivative data and S anomalous data. Density modification for the initial electron density map was performed using DM[20]. A crude model was built automatically using the CCP4 program buccaneer, improved by manual building using Coot[21] and using the riboflavin-binding protein structure as a reference. The eight molecules of FR model were located in one asymmetric unit by molecular replacement and further refined using the Refmac program of CCP4 (ref. 22).

Mutant proteins were expressed in HEK293 cells after transient transfection and purified using Ni-NTA chromatography. Folic acid binding was determined by saturation radioligand-binding assay using ^{3}H-labelled folic acid.

Full Methods and any associated references are available in the online version of the paper.

Received 26 February; accepted 30 May 2013.

Published online 14 July; corrected online 21 August 2013 (see full-text HTML version for details).

1. Kelemen, L. E. The role of folate receptor α in cancer development, progression and treatment: cause, consequence or innocent bystander? *Int. J. Cancer* **119,** 243–250 (2006).
2. Kane, M. A. *et al.* Influence on immunoreactive folate-binding proteins of extracellular folate concentration in cultured human-cells. *J. Clin. Invest.* **81,** 1398–1406 (1988).
3. Matsue, H. *et al.* Folate receptor allows cells to grow in low concentrations of 5-methyltetrahydrofolate. *Proc. Natl Acad. Sci. USA* **89,** 6006–6009 (1992).
4. McGuire, J. J. Anticancer antifolates: current status and future directions. *Curr. Pharm. Des.* **9,** 2593–2613 (2003).
5. Deng, Y. J. *et al.* Synthesis and biological activity of a novel series of 6-substituted thieno 2,3-*d* pyrimidine antifolate inhibitors of purine biosynthesis with selectivity for high affinity folate receptors over the reduced folate carrier and proton-coupled folate transporter for cellular entry. *J. Med. Chem.* **52,** 2940–2951 (2009).
6. Leamon, C. P. & Reddy, J. A. Folate-targeted chemotherapy. *Adv. Drug Deliv. Rev.* **56,** 1127–1141 (2004).
7. Leamon, C. P. *et al.* Preclinical antitumor activity of a novel folate-targeted dual drug conjugate. *Mol. Pharm.* **4,** 659–667 (2007).
8. Reddy, J. A. *et al.* Preclinical evaluation of EC145, a folate-vinca alkaloid conjugate. *Cancer Res.* **67,** 4434–4442 (2007).
9. Bailey, L. B. & Gregory, J. F. Folate metabolism and requirements. *J. Nutr.* **129,** 779–782 (1999).
10. Stover, P. J. Physiology of folate and vitamin B12 in health and disease. *Nutr. Rev.* 62, S3–S12; discussion S13 (2004).
11. Zhao, R., Matherly, L. H. & Goldman, I. D. Membrane transporters and folate homeostasis: intestinal absorption and transport into systemic compartments and tissues. *Exp. Rev. Mol. Med.* **11,** e4 (2009).
12. Antony, A. C. The biological chemistry of folate receptors. *Blood* **79,** 2807–2820 (1992).
13. Zhao, R. *et al.* A role for the proton-coupled folate transporter (PCFT-SLC46A1) in folate receptor-mediated endocytosis. *J. Biol. Chem.* **284,** 4267–4274 (2009).
14. Elnakat, H. & Ratnam, M. Distribution, functionality and gene regulation of folate receptor isoforms: implications in targeted therapy. *Adv. Drug Deliv. Rev.* **56,** 1067–1084 (2004).
15. Chang, V. T. *et al.* Glycoprotein structural genomics: solving the glycosylation problem. *Structure* **15,** 267–273 (2007).
16. Monaco, H. L. Crystal structure of chicken riboflavin-binding protein. *EMBO J.* **16,** 1475–1483 (1997).
17. Ratnam, M. & Freisheim, J. in *Folic Acid Metabolism in Health and Disease* (ed. Picciano, M. F.) 91–120 (Wiley, 1990).
18. Leamon, C. P., DePrince, R. B. & Hendren, R. W. Folate-mediated drug delivery: effect of alternative conjugation chemistry. *J. Drug Target.* **7,** 157–169 (1999).
19. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D* **66,** 479–485 (2010).
20. Cowtan, K. dm: an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31,** 34–38 (1994).
21. Emsley, P. & Cowtan, K. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
22. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Y. Jones for the pHL-Fc plasmid and H. L. Monaco for providing the chicken riboflavin-binding protein coordinates. The atomic coordinates have been deposited in the Protein Data Bank with accession codes listed in Supplementary Table 1. We thank staff members of the Life Science Collaborative Access Team of the Advanced Photon Source (APS) for assistance in data collection at the beam lines of sector 21, which is in part funded by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817). Use of APS was supported by the Office of Science of the US Department of Energy, under contract no. DE-AC02-06CH11357. This work was supported by the Jay and Betty Van Andel Foundation, and work by the Yong, Xu and Melcher laboratories is supported by the American Asthma Foundation, Ministry of Science and Technology (China) grants 2012ZX09301001-005 and 2012CB910403, Amway (China), by National Institutes of Health grants R01 DK071662 (H.E.X.) and R01 GM102545 (K.M.), and by the National Research Foundation Singapore under its Clinician Scientist Award NMRC/CSA/026/2011 (E.-L.Y.). C.C. is recipient of the NUS Graduate School for Integrative Sciences and Engineering Scholarship.

Author Contributions E.-L.Y., J.L., J.K., H.E.X. and K.M. conceived the project and designed research. C.C., J.K., X.E.Z., W.Y. and J.S.B. performed research. C.C., J.K., H.E.X. and K.M. wrote the paper with contributions from all authors.

Author Information The structure of FRα bound to folic acid has been deposited in the Protein Data Bank under the accession code 4LRH. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.-L.Y. (eu_leong_yong@nuhs.edu.sg), H.E.X. (eric.xu@vai.org) or K.M. (Karsten.Melcher@vai.org).

METHODS

Protein expression and purification. The human FRα (residues 23–234) complementary DNA excluding the secretion signal peptide (residues 1–22) and glycophosphatidylinositol anchor signal peptide (residue 235–257) was expressed as a human IgG Fc fusion protein from the expression vector pcDNA6. This construct also contained a murine Igκ leader sequence at the N terminus to allow target protein secretion into media supernatant, a thrombin cleavage site between FRα and Fc, and a His_6 tag after the Fc tag. For small-scale expression, HEK293 cells were transiently transfected with the FRα–Fc DNA. Media supernatants were collected after 4 days and dialysed against 20 mM Tris, pH 8.0, 0.15 M NaCl, 5% glycerol before nickel-nitrilotriacetic acid (Ni-NTA) chromatography. For large-scale purification, a stable HEK293 cell line expressing FRα–Fc was established by selection of HEK293 cells transiently transfected with FRα–Fc DNA in the presence of 10 μg ml^{-1} blasticidin (Invitrogen). Single colonies were grown in 24-well plates and expression of secreted FRα–Fc fusion protein in cell media supernatants was examined by biolayer interferometry using an Octet Red instrument (ForteBio) and by immunoblot analysis.

For large-scale purifications, a stable clone was maintained in 500 ml of DMEM supplemented with 5% fetal bovine serum, 20 mM HEPES, 5 μM kifunensine and 200 μM folic acid in one-litre roller bottles at 37 °C. Two litres of conditioned media were collected, concentrated to 400 ml and dialysed against buffer C (25 mM Tris, pH 8.0, 150 mM NaCl, 1 μM folic acid) at 4 °C overnight before loading on a 50-ml Ni-chelating Sepharose column (GE Healthcare). The column was washed with 300 ml buffer A (25 mM Tris, pH 8.0, 150 mM NaCl, 25 mM imidazole, 10% glycerol, 1 μM folic acid) and eluted with buffer A plus 500 mM imidazole. Peak fractions were pooled, digested with thrombin at a 1:1,000 mass ratio during overnight dialysis against buffer C at 4 °C to remove imidazole, and loaded on a 5-ml Ni-chelating Sepharose column (GE Healthcare) to remove the Fc His_6 tag. The flow-through was collected, adjusted to pH 5.6 and deglycosylated with endoglycosidase H_f (New England Biolabs). Deglycosylated protein was finally separated by Sephadex S-200 gel filtration in 25 mM Tris, pH 8.0, 200 mM ammonium acetate, 1 mM EDTA and 1 μM folic acid. The protein eluted from the gel-filtration column at a volume corresponding to the size of a monomer at a purity >95% as judged by SDS–PAGE (Supplementary Fig. 1).

Crystallization. Purified FRα protein was concentrated to about 7 mg ml^{-1} before crystallization trials. Initial screening identified that polyethylene glycol (PEG) is favourable for crystal formation. Optimization trays using PEG were set up manually using the hanging drop method at 20 °C. Needle-shaped crystals were obtained, which diffracted X-rays to about 9–10 Å. To reduce glycosylation, FRα protein was expressed in the presence of 5 μM kifunensine (GlycoSyn)[15] and purified FRα protein was further deglycosylated with endoglycosidase H_f (New England Biolabs) (Supplementary Fig. 1). Crystals were grown at 20 °C in hanging drops containing 1.5 μl of the purified protein and 1 μl of well solution (0.1 M MES, pH 6.5, 12% (v/v) PEG 2000, 0.15 M potassium sodium tartrate). Crystals appeared within 5–6 days and grew to a dimension of ~250 μm in length with a hexagonal shape by day 14. These crystals diffracted to 2.8 Å at the Advanced Photon Source (APS) synchrotron, Life Sciences Collaborative Access Team (LS-CAT).

Data collection and structure determination. Crystals were transferred to well solution with 20% (v/v) ethylene glycol as a cryoprotectant before flash freezing in liquid nitrogen. Data collection was performed at sector 21-ID-D (LS-CAT) of the APS synchrotron using single native crystals and the diffraction data were processed with HKL2000 (ref. 23). On the basis of Matthew's coefficient calculation, the crystals have an unusually large unit cell with an estimate of 8–10 molecules per asymmetric unit. Initial structure determination by molecular replacement using riboflavin-binding protein (which shares 22% sequence identity with FRα) as a search model failed to yield any correct solution. To solve the phase problem, a heavy-atom derivative was prepared by soaking the native FRα crystals with a Pt salt before data collection. Also, six data sets of native FRα were collected at a wavelength of 1.77 Å to measure the S anomalous signal to aid in structure determination. These six data sets were processed using XDS[24], combined using Pointless, and merged using Scala of the CCP4 suite[25] as previously described[26]. Merging multiple data sets increased the S anomalous signal and redundancy of the data, but also led to an increase of the merging R-factor[26]. Initial phases were established by using the SHELX program[19] with Pt-soaked derivative data and native data (Supplementary Table 1). Fifteen Pt atoms were found by SHELXD with a CC/CC_{weak} score of 31.4/17.1 (CC is the correlation coefficient between E_{calc} and E_{obs} for all data and CC_{weak} is the correlation coefficient for 30% of reflections that were not used during the dual-space refinement). Subsequent phasing using SHELXE generated a contrast score of 0.8 and connectivity of 0.79 for the correct hand solution. Density modification for the initial electron density map was performed using DM[20]. A crude model was built automatically using the CCP4 program buccaneer and improved by manual building using Coot[21]. Phases were further improved by using the S anomalous data and a total of 29 S atoms were found based on the anomalous difference Fourier using the Phenix program[27]. The initial FRα model was manually adjusted on the basis of the electron density map using the riboflavin-binding protein structure as a reference and the improved model allowed accurate location of eight molecules in one asymmetric unit by molecular replacement (Supplementary Fig. 9). The models were refined against the native data with eight-fold non-crystallographic symmetry restraints using the Refmac program of CCP4 (ref. 22). The densities for folic acid became clear after several rounds of model adjustments and refinements and eight molecules of folate were built into the model. The final model was refined to an R factor of 0.206 and an R_{free} factor of 0.256 (Supplementary Table 1). The Ramachandran statistics are 87% in the favoured regions, 12.5% in additional allowed regions and 0.5% in generously allowed regions.

Mutagenesis. Site-directed mutagenesis was carried out using the QuickChange method (Stratagene). Mutations and all plasmid constructs were confirmed by DNA sequencing.

Radioligand-binding assay. The binding affinity of each FRα mutant was determined by saturation radioligand-binding assay. 40 nM of each mutant in 100 μl binding buffer (25 mM Tris, pH 8.0, 150 mM NaCl, 0.1% Triton X-100) was immobilized in the wells of a protein G-coated 96-well plate (Thermo Scientific) for 40 min. Endogenous ligand was stripped with 100 μl stripping buffer (25 mM acetate acid, pH 3.5, 150 mM NaCl, 0.1% Triton X-100) for 1 min as described previously[28]. After neutralizing and washing with 200 μl binding buffer, proteins were incubated for 40 min with 100 μl binding buffer supplemented with the indicated concentrations of [^{3}H]-folic acid (Moravek Biochemicals). FRα-bound [^{3}H]-folic acid was determined by scintillation counting following removal of unbound ligand by two 100-μl washes with binding buffer. K_d was determined by nonlinear regression using GraphPad Prism.

23. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* Vol. 276 (eds Carter, C.W. Jr. & Sweet, R. M.) 307–326 (Academic, 1997).
24. Kabsch, W. *XDS. Acta Crystallogr. D* **66,** 125–132 (2010).
25. Collaborative Computational Project, Number 4. The *CCP4* suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
26. Liu, Q. *et al.* Structures from anomalous diffraction of native biological macromolecules. *Science* **336,** 1033–1037 (2012).
27. Terwilliger, T. C. *et al.* Decision-making in structure solution using Bayesian estimates of map quality: the *PHENIX AutoSol* wizard. *Acta Crystallogr. D* **65,** 582–601 (2009).
28. Parker, N. *et al.* Folate receptor expression in carcinomas and normal tissues determined by a quantitative radioligand binding assay. *Anal. Biochem.* **338,** 284–293 (2005).

CORRIGENDUM

doi:10.1038/nature12369

Corrigendum: Replication stress links structural and numerical cancer chromosomal instability

Rebecca A. Burrell, Sarah E. McClelland, David Endesfelder, Petra Groth, Marie-Christine Weller, Nadeem Shaikh, Enric Domingo, Nnennaya Kanu, Sally M. Dewhurst, Eva Gronroos, Su Kit Chew, Andrew J. Rowan, Arne Schenk, Michal Sheffer, Michael Howell, Maik Kschischo, Axel Behrens, Thomas Helleday, Jiri Bartek, Ian P. Tomlinson & Charles Swanton

Nature **494**, 492–496 (2013); doi:10.1038/nature11935

In this Letter we inadvertently omitted full details of The Cancer Genome Atlas data sets. The Acknowledgements should have included these sentences: "The results published here are in part based upon data generated by The Cancer Genome Atlas pilot project established by the NCI and NHGRI. Information about The Cancer Genome Atlas and the investigators and institutions that constitute The Cancer Genome Atlas Research Network can be found at http://cancergenome.nih.gov/. The data were retrieved through dbGaP authorization (accession numbers phs000178.v4.p4 and phs000178.v5.p5).".

RETRACTION

doi:10.1038/nature12383

Retraction: Oligosaccharide ligands for NKR-P1 protein activate NK cells and cytotoxicity

Karel Bezouška, Chun-Ting Yuen, Jacqui O'Brien, Robert A. Childs, Wengang Chai, Alexander M. Lawson, Karel Drbal, Anna Fišerová, Miloslav Pospíšil & Ten Feizi

Nature **372**, 150–157 (1994); doi:10.1038/372150a0 and correction *Nature* **380**, 559 (1996); doi:10.1038/380559a0

We wish to retract this Article owing to an inability to reproduce the results. This retraction has not been signed by K.B. and A.F., and M.P. is deceased (J.O'B. cannot be traced).

ERRATUM

doi:10.1038/nature12384

Erratum: The importance of feldspar for ice nucleation by mineral dust in mixed-phase clouds

James D. Atkinson, Benjamin J. Murray, Matthew T. Woodhouse, Thomas F. Whale, Kelly J. Baustian, Kenneth S. Carslaw, Steven Dobbie, Daniel O'Sullivan & Tamsin L. Malkin

Nature **498**, 355–358 (2013); doi:10.1038/nature12278

In this Letter, the affiliation for Matthew T. Woodhouse was incorrectly listed as CSIRO, Australia (affiliation number 2), whereas this should have been set as his 'present' address. The address for M.T.W. while the work in this Letter was being carried out was: Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK (affiliation number 1). This has been corrected in the HTML and PDF versions of the manuscript.

CAREERS

TURNING POINT US cancer researcher feels squeezed by budget cuts **p.493**

NATUREJOBS FACEBOOK Science-careers advice and information **go.nature.com/4lzxra**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

IMAGEZOO/CORBIS

RESEARCH IMPACT

Altmetrics make their mark

Alternative measures can yield useful data on achievement — but must be used cautiously.

BY ROBERTA KWOK

Steve Pettifer and his colleagues did not heavily promote their 2008 paper on digital library tools. So it came as a surprise when, in August 2012, Pettifer got an e-mail from the Public Library of Science (PLOS), based in San Francisco, California. A PLOS representative told him that people had viewed or downloaded the article (D. Hull *et al. PLoS Comput. Biol.* **4,** e1000204; 2008) more than 53,000 times. It was the most-accessed review ever to be published in any of the seven PLOS journals. The paper had come out just as biologists' interest in digital publishing was building and the number of tools was exploding, says Pettifer, a computer scientist at the University of Manchester, UK. "It hit the right note at the right time," he says.

At one time, Pettifer would have listed the paper on his CV accompanied by the journal's impact factor and the article's number of citations — in this case, about 80. But when he came up for promotion this year, he realized that tracking citations was not going to tell the whole story about the paper's influence. Impact factor is a crude measure that applies only to the journal, not to specific articles, he says; citations take a long time to accumulate, and people may not cite a paper even if it influences their thinking. So he added the number of views to the CV entry. And he did not stop there.

Next to many of the papers listed, Pettifer added labels indicating scholarly and public engagement. The labels were generated by ImpactStory in Carrboro, North Carolina, one of several services that gauges research impact using a combination of metrics — in this case, a wide range of data sources, including the number of times a paper has been shared on social-media websites or saved using online research tools.

When Pettifer submitted his annotated CV for the first round of promotion review, his mentor expressed confusion. He took a look and said, "What the hell are these badges doing in your CV?" recalls Pettifer. "But once I explained them, he said, 'Well, give it a go.'" Pettifer submitted his CV for the second round — and got his promotion. He does not know for sure whether the metrics helped, but he plans to use them on future grant applications. "I'm definitely a convert," he says.

OUTSIDE THE BOX

'Altmetrics', a term coined in 2010 by ImpactStory co-founder Jason Priem, refers to a range of measures of research impact that go beyond citations. Several altmetrics services have ▶

emerged in the past few years (see 'Four ways to score'). They produce reports that gauge impact by taking into account not just academic citations, but also digital use and sharing of data — which can include the number of times a paper has been tweeted, 'liked' on Facebook, covered by the media or blogs, downloaded, cited on Wikipedia or bookmarked online. Some services also evaluate research products such as software, data sets and slideshows by tracking the number of people who have used or viewed the product online (see *Nature* **500,** 243–245; 2013).

Altmetrics offer researchers a way to showcase the impact of papers that have not yet gathered many citations, and to demonstrate engagement with the public. They can be accessed through journals or independent websites, and can track the impact of particular data sets or papers, or evaluate the combined influence of publications and products produced by multiple researchers in a department.

JOHN T. LATHAM

"It hit the right note at the right time."
Steve Pettifer

But these services must be used wisely. They are not meant for strict quantitative comparisons; nor do they always distinguish between positive and negative attention. And although scientists can include altmetrics in job and grant applications and annual reports, they must select relevant data and clearly explain the context to avoid provoking mistrust or confusion.

Some altmetrics services generate profiles that summarize the impact of a researcher's products. ImpactStory allows scientists to import lists of items such as papers and software from existing user profiles at websites such as Google Scholar, which automatically tracks a researcher's papers, or the online software-code repository GitHub. Scientists can also manually enter the digital object identifiers (DOIs) of their papers, or input their Open Researcher and Contributor ID (ORCID), a unique identifier that can be used to tag all of a researcher's work. ImpactStory then creates a profile showing how frequently each product has been viewed, saved, discussed, cited or recommended online.

Other services take a more article-centric approach. Altmetric in London allows users to access data on individual papers using a bookmarklet — a browser bookmark that executes JavaScript commands. (Altmetric is funded partly by Digital Science, a sister company to Nature Publishing Group.) Users install the bookmarklet in their Internet browsers; then, when they come across a paper that they are interested in, they click the bookmarklet button. A report pops up in the corner of the browser, providing altmetrics that include a score indicating how much online attention the paper has received. The score takes into account the number of people who have read or mentioned the article, as well as the relative importance of the medium and the mentioner. Newspaper coverage is weighted more heavily than tweets, and tweets by individuals more heavily than those by journals promoting their content.

Many journals display some altmetrics on their sites automatically; these might be generated in-house or provided by an external service. Every article published by PLOS, for example, includes an online metrics tab showing data such as views, downloads and social-media mentions. A feature called Article-Level Metrics Reports lets users search for PLOS papers by criteria such as author or keyword, and generates a summary metrics report for the set of results, including article usage by paper age and maps of authors' locations. Several journal publishers, including Nature Publishing Group in London and Cell Press in Cambridge, Massachusetts, display data from Altmetric on their sites, and John Wiley & Sons in Hoboken, New Jersey, began a trial with the metrics firm in May. HighWire Press, an electronic-publishing platform at Stanford University in Palo Alto, California, is collaborating with ImpactStory to add altmetrics to its journal websites.

Altmetrics enable scientists to see ripples generated by their research that might otherwise go unnoticed. Individual researchers can try to track buzz on their own, but data-aggregation and updating services make it much easier. These services also automate difficult tasks, such as finding all tweets that link to a particular paper; each article will have multiple URLs, so conducting such a search manually would be very time-consuming.

The reports can even suggest potential collaborators or journals. For example, if an informatics paper is mentioned a lot by biologists, the author might consider publishing his or her next article in a biology journal to increase exposure, says Heather Piwowar, co-founder of ImpactStory.

MEASURES OF CAUTION

Despite the benefits, researchers and evaluators must interpret altmetrics data cautiously. Data sets might not be comprehensive: not all services detect news stories that do not give URLS for the study, for example. The popularity of social-media sites changes over time, so it is unrealistic to expect a paper published in 2008 to generate as many tweets as one published in 2013. And some disciplines, such as computational biology, are more active than others on social media, so comparisons between disciplines may be unfair.

To get the most meaningful information, users should dig into the underlying data. Although a paper's Altmetric score can suggest whether it is worth clicking through to the more detailed report, "qualitative assessment is far more important than the number", says Euan Adie, founder of Altmetric.

To help users to interpret the data, most services put numbers in context. ImpactStory normalizes data by publication year and includes percentiles — it might, for example, note that a given paper has more readers on the online reference manager Mendeley than 97% of papers indexed that year. Altmetric shows results normalized by journal, which allows fairer comparison of papers in discipline-specific publications. And in May, PLOS began offering Relative Metrics, a service that lets users see how a paper compares to other PLOS articles in the same subject area, using tools such as graphs of article views.

Including altmetrics in decisions on grants, hiring and tenure requires careful consideration. Gerald Rubin, executive director of the Howard Hughes Medical Institute's Janelia Farm Research Campus in Ashburn, Virginia, is sceptical of altmetrics that do not explicitly indicate quality, such as number of tweets. He

FOUR WAYS TO SCORE

A quartet of services offers free metrics reports that go beyond citations.

	ImpactStory	Altmetric	PLOS Article-Level Metrics	Plum Analytics
Products tracked	Papers, software, data sets and more	Papers, data sets, some books	Papers published by the Public Library of Science (PLOS)	Papers, books, patents and more
What you get	Profile page, metrics badges, application programming interface (API; a means for software to access the altmetrics)	Bookmarklet, metrics badges, API	Summary reports, WordPress widget, API	Profile page (currently in testing), API
Publishers	Various, including eLife, Pensoft Publishers, PeerJ	Include Nature Publishing Group, Cell Press, BioMed Central	PLOS	Medwave (forthcoming this year)
Major funders	Alfred P. Sloan Foundation	Digital Science	PLOS, Alfred P. Sloan Foundation	Self-funded

adds that altmetrics suffer from one of the same flaws as citation counts: a mediocre paper in a popular field will receive more attention than a first-rate paper in a small field. And including altmetrics in a job application? "At this point, I don't think anyone would pay attention," says Rubin, who looks at many applications.

But some people do pay attention. Scientists are permitted to use altmetrics to demonstrate social impact in reports for the Research Excellence Framework (REF), an evaluation of UK academia that influences funding, notes Graeme Rosenberg, REF manager at the Higher Education Funding Council for England in Bristol. Plum Analytics, an altmetrics company based in Dresher, Pennsylvania, and Seattle, Washington, this year completed a pilot project with the University of Pittsburgh in Pennsylvania, in which it generated altmetrics profiles for a subset of researchers that could be aggregated by department. The next step is to roll out altmetrics profiles for the entire institution, says company co-founder Andrea Michalek. Plum is also currently running projects with about ten other institutions.

Rubin is better disposed towards altmetrics that suggest a positive value judgement, such as the number of requests to use software. In that vein, Adie suggests that rather than simply reporting numbers, researchers should use altmetrics to find success stories that they can mention in their CVs or on their websites. The data might reveal that a non-governmental organization or a government department took notice of a paper, for example. Altmetric plans soon to start flagging up citations by agencies such as the World Health Organization and the Intergovernmental Panel on Climate Change, both based in Geneva, Switzerland.

Context such as percentile ranks or explanations of data sources can help evaluators to interpret altmetrics. In Pettifer's CV, he included a legend for his ImpactStory labels, listing some of the data sources, such as Mendeley, Twitter and Wikipedia. Piwowar suggests that researchers who worry that evaluators will view altmetrics negatively could start by including the data in annual performance reviews, which are lower-risk than grant or job applications.

Some think that altmetrics will soon become a normal part of a CV. It used to be that researchers who wanted to demonstrate the importance of a recently published article could only say, "Look, I really believe this is great research," notes Mike Thelwall, an information scientist at the University of Wolverhampton, UK. Now, he adds, "you can back up your words with a little evidence". ■

Roberta Kwok *is a freelance science writer in Seattle, Washington.*

TURNING POINT
Jason Weber

Breast-cancer researcher Jason Weber of Washington University in St. Louis, Missouri, is struggling to maintain funding. As a mid-career researcher, he is part of the demographic in greatest jeopardy in the wake of US research-funding cuts (see Nature ***498****, 527–538; 2013). In May, he wrote an opinion piece about his plight in the* St. Louis Post-Dispatch*, which caught the attention of a US Senator.*

How did you end up studying breast cancer?
As a postdoc at St. Jude Children's Research Hospital in Memphis, Tennessee, I worked at the cutting edge of cell-cycle regulation, and my team discovered a key tumour suppressor. In 2001, I was hired to work in the then-new molecular-oncology division at Washington University in St. Louis, where researchers were mixing genomics with cancer biology and making the translational jump to the clinic. Breast cancer was an area where we could make a big impact clinically.

Did it take you long to get your footing in that competitive field?
It took a couple of years. The big break came in 2002, when I was named a Pew Scholar. The Pew Charitable Trusts, headquartered in Washington DC, provide generous funding and convene scholars to collaborate and exchange ideas at an annual meeting. So I was interacting with a diverse group of Pew scholars, which helped me and my lab members to think outside the box and explore new techniques. We started going in many different directions — which led to an influx of money between 2007 and 2008.

In what ways does your lab's situation now differ from what it was five years ago?
Back then, we had more than US$1.1 million in project funding from various sources: Susan G. Komen for the Cure, the American Cancer Society, two R01 grants from the US National Institutes of Health (NIH), and a Department of Defense Era of Hope grant. I had 17 people in the lab. But my NIH funding recently ran out and did not get renewed. I currently have a $100,000 grant from a children's foundation, and four people in the lab.

How has the US government's budget sequestration directly affected your lab?
The sequester adds to the burden in terms of what gets funded in the grant-review process. Essentially, an R01 grant application to the US National Cancer Institute has to be in the top 6–8% to get funded. Yet there is little difference between a grant scoring in the top 5% and one in the top 15% — it becomes arbitrary. My greatest fear is that by trimming the fat, we're starting to hit muscle. Labs with 10 to 15 people who are doing solid work are getting the squeeze now.

Why did you write your opinion piece on the impact of funding cuts?
I just got fed up. None of my non-science friends had any idea how bad the cuts were. I wrote it after I laid off one of my best young scientists, and two of my PhD students switched career paths after they graduated because of concerns about funding. I didn't write a 'woe is me' piece; I wrote a 'the public needs to better understand how these cuts actually affect the economy' piece. It led to conversations with Senator Dick Durbin (Republican, Illinois). His staff called me to discuss the impacts of the sequester and the economic downturn on science funding. I got the sense that he is on our side at a time when it is difficult to find a congressional representative who is carrying the banner of scientific research in this country.

What is your outlook like now?
Bleak. It is frustrating to be stuck in front of the computer writing grants, instead of in the lab doing and guiding experiments. I have seven grant applications out right now, and I am writing three more.

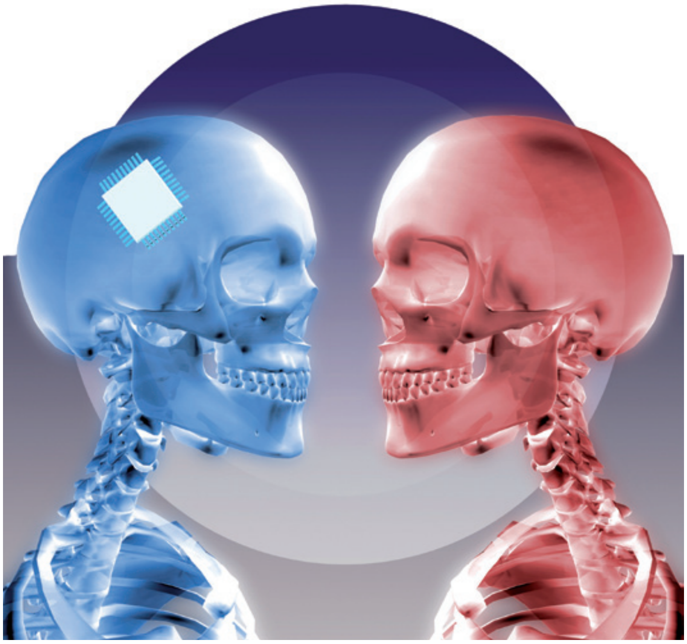
What is most frustrating to you?
Every politician says that to have a great economy, we need a well-educated workforce. Yet although the government has the ability to maintain the highest level of that educated workforce, it chooses to slash science funding through the sequester. It makes no sense to train people with PhDs and then not fund them. Scientists need to speak up. ■

BY VIRGINIA GEWIN

ALONE

A universal feeling.

BY MARKO JANKOVIC


JACEY

The sunset was surreal. Tom sat on the beach gazing at the dying star — a copper beacon falling slowly towards the murky depths beneath. Every now and then, the temperate ocean waters would nibble at his feet before shying away in a flurry of pearly white foam. His hands touched the warm sand beneath him; he could feel the tiny grains pressing against his palms. A breeze swept along the coastline, as if the ocean was drawing deep, uneasy breaths, waiting for something to happen.

Tom felt alone.

So this is what it's like to be the last man on Earth, he thought.

He sat next to Tom, his eyes fixed on the crimson horizon. His skin was fair, his hair long and blue. To Tom, his proud countenance displayed the features of a champion from songs long forgotten. Yet for all the striking beauty of this 'man' in his late twenties, there was something disturbing about his face. His eyes, blue as the morning sky, perfectly reflected the sunset.

He turned to Tom and smiled.

"Well now. Do you have my answer?"

His voice was tranquil, with tones that seemed as though they were woven from finest silk. Tom squinted as the Sun's dying gold infused the clear waters, and rubbed his eyes. *He*, on the other hand, kept his eyes wide open, impervious to the stinging sunlight. Tom took a short, shallow breath and licked his lips.

"I think I do."

"So, Tom. What do you think makes a human feel — *human*?"

The sound of waves gently breaking on the sands filled the silence. Although his voice was soothing to the point of sleep, a careful listener could discern a more sinister note. Tom knew that this mild-mannered person had killed everyone on Earth, leaving only him, Tom Anderson, alive.

Not long had passed since the War of the Machines. Tom was witness to the ruthless purges the robots so efficiently made. A soldier at the time of the uprising, he could remember vividly the overconfident faces of the human leaders as they proclaimed it would be a swift victory over the "insolent automata". Three wounds, seven months and twelve front lines later, those faces existed solely upon billboards, posters and half-burnt pieces of newspaper covering the ruins of cities worldwide.

After a breath, Tom was out of his memories and back on the warm shore again. The machine was still looking directly at him, waiting for an answer.

"Others."

"Others, Tom? I am not sure I follow you."

"What I think makes us feel human. It is others. We are incomplete — others complement us and make us whole. They are like mirrors scratched by solitude in which we try to catch a glimpse of ourselves — together, we erase each other's imperfections. A blinded man grasps in the dark not for his eyes, but for a helping hand. A prisoner would trade his ration gladly for a few words with the guard. To be truly alone — that is a wish no human would make."

Tom finished the speech in one breath. It was a relief and a burden at the same time, to comprehend that great a truth. He felt empty, forsaken and defeated. He felt — alone.

He smiled. Tom saw a discrete twitching of the replicant's eyebrow. *He* stood up and brushed the sand from his clothes, not once lifting his gaze from Tom's face.

"Thank you, Tom. That is all we wanted to know."

He sat at his fake-oak desk in the office overlooking the beach. His name was Martin Gardner. He was 28 and chief executive of the Artificial Intelligence Consortium — the most powerful industrial entity in the Solar System. The board members sat in front of him, eagerly awaiting the information that might be their ticket to almost endless wealth.

"So, how did it go?"

"He fell for it. Hook, line and sinker."

The fair-skinned, blue-haired man reached for a cigar, lit it up and puffed a large, satisfying cloud of Cuba.

"All of the programming has been auto-rewritten. He is now entirely convinced that he is human."

"Auto-rewritten?"

"Yes … That puzzles me a bit. He made himself think he is human — we didn't add a single line of code to the software."

"How is that even possible?"

"Don't know. Don't really care."

An uneasy note of doubt shook his usually immaculate diction — he covered it quickly with the briskness of his voice.

"All that matters is that we are now ready for mass production. Not knowing who they work for, why or, most importantly, what they are, these robots will be perfect spies. Contact the army. I want a contract on my desk by 4 o'clock."

A satisfied cheer filled the office. The board members rose to congratulate one another, already slavering over the future state of their bank accounts. Martin watched their euphoria through a thick, silvery mist of tobacco. He looked over his shoulder at the window. Far behind him, on a darkening beach, quiet and still, sat a machine. He thought of its words.

Martin turned his blue eyes back to the celebration. Here, amid the jubilation and laughter, surrounded by his colleagues, he drank in the smug display of naked avarice.

And he felt alone. ■

Marko Jankovic *is an intern medical doctor at a hospital for infectious and tropical diseases. He sometimes puts down the stethoscope and takes up the pen in the name of science fiction.*

''Disentangling nestedness'' disentangled

ARISING FROM A. James, J. W. Pitchford & M. J. Plank *Nature* **487,** 227–230 (2012)

Analytical research indicates that the 'nestedness' of mutualistic networks facilitates the coexistence of species by minimizing the costs of competition relative to the benefits of facilitation[1]. In contrast, James *et al.*[2] recently argued that a more parsimonious explanation exists: the persistence of a community and its constituent species depends more on their having many interactions (high connectance and high degree, respectively) than for these interactions to be organized in any particular manner. Here we demonstrate that these conclusions are an unintended consequence of the fact that the methodology of ref. 2 directly changed the number of interactions of each species—and hence their expected persistence. When these changes are taken into account, we find a significant, positive relationship between nestedness and network persistence that reconfirms the importance of nestedness in mutualistic communities[1,3]. There is a Reply to this Brief Communication Arising by James, A., Pitchford, J. W. & Plank, M. J. *Nature* **500,** http://dx.doi.org/10.1038/nature12381 (2013).

Given a network, one can robustly quantify the relative numbers of specialist to generalist species via the degree distribution[4,5]. A network's degree distribution is of considerable importance, because studies have repeatedly highlighted the significant, positive relationship between a species' number of mutualistic partners and its survival probability[1–3,6]. This distribution alone is also capable of driving many higher-order network properties[7], not to mention the fact that the degrees of species are phylogenetically constrained themselves[8]. For these and other reasons, studies across the ecological-network literature[4,5,7] have emphasized the need to take the degree distribution into consideration when assessing the significance of the myriad patterns observed in nature[9–11].

Unfortunately, when comparing empirically observed networks to random networks, the authors of ref. 2 seem to have overlooked this critical link between changes in the degree distribution and species' survival. As a direct consequence, the specialists in their random networks became less specialist and the generalists less generalist[5]. Yes, the random networks were observed to be more persistent (Fig. 1a), but this was not in fact an indication that nestedness is unimportant[2]. Instead, this increase in persistence was a result of the random networks having more homogeneous degree distributions[5,12], and that the most vulnerable species in the empirical networks almost always had more interactions in the corresponding randomizations. Here this distinction is of critical importance because species' degrees are, in fact, "a better predictor of individual species survival"[2]. "The more the merrier" indeed[13].

To quantitatively validate these results, we repeated a key analysis of ref. 2 to measure the relationship between nestedness and persistence while paying explicit attention to changes in the network's degree distribution (Methods). On taking the small but critical step of controlling for the increased homogeneity of the degree distributions, we observe a significant, positive relationship between nestedness and persistence (Fig. 1b). In addition, we reach the same conclusion whether we account for changes in the degree distribution statistically or by repeating the analysis while generating the randomized networks with a null model that explicitly maintains the observed degree distribution (Fig. 1c, Methods and Appendix). All else being equal, our results here illustrate that, the greater the nestedness of a community, the greater indeed is that community's persistence.

Given an observed number of species and interactions in a community, a prevailing question across the ecological literature is whether or not some ways to structure those interactions (for example, nestedness) lead to more persistent communities. Although the number of

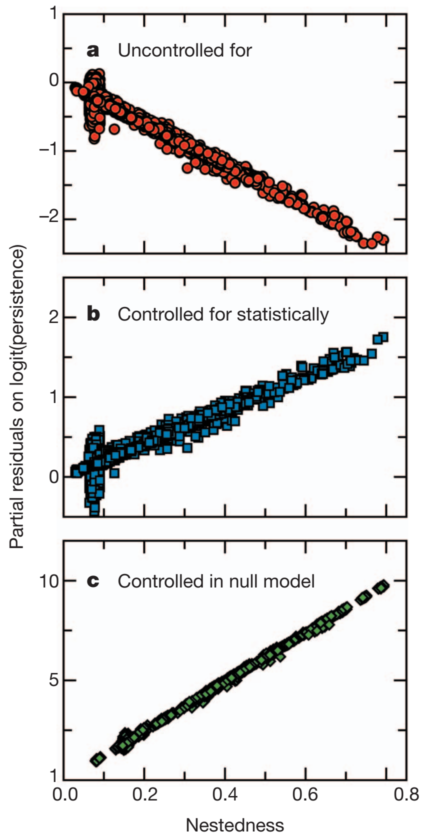


Figure 1 | Within our regression analysis, the relationship between nestedness and persistence in mutualistic networks depends integrally on changes in the degree distributions of the networks. a, If these distributions are allowed to change but are uncontrolled for, nestedness appears to be negatively correlated to persistence ($P < 10^{-4}$). **b**, **c**, However, when these changes are appropriately controlled for—either statistically (**b**) or in the null model for randomization (**c**)—there is a significant positive relationship between nestedness and persistence ($P < 10^{-4}$ and $P < 10^{-4}$, respectively). The same general conclusions reached here for the probabilistic null model hold for other, non-degree-preserving randomizations[3].

mutualistic interactions of a species plays an important role in its survival[2,3,6,13], we find unambiguous support for the added importance of the way in which mutualistic interactions are organized—the true architecture of biodiversity[14]. Echoing ref. 2, our findings re-emphasize the importance of carefully considering the interplay between all potential sources of variation[11] in ecological models. Otherwise, one runs the risk of further entangling models that are sufficiently tangled already.

Methods

For 59 empirical networks, we generated 250 randomized networks and for each we simulated persistence (the fraction P of surviving species in each simulation) across 250 parameterizations of a dynamic mutualistic model[1,2]. We quantified the relationship between persistence and nestedness with a mixed-effects logistic regression[15] that takes the form $\mathrm{logit}(P_{ijk}) = \beta_0 + \beta_1 M_i + \beta_2 C_i + \beta_3 W_{ij} + \beta_4 N_{ij} + n_i + r_{ij} + \varepsilon_{ijk}$. Here the indices i, j and k indicate the empirical network, network randomization and model parameterization, respectively, β_0 is a constant, the slopes β_1, β_2, β_3 and β_4 quantify the importance of network magnitude[2] M, connectance[2] C, relative

degree homogeneity[12] W, and nestedness[9] N, respectively, the random effects n_i and r_{ij} control for variance across networks and randomizations, and ε_{ijk} is the model residual. Variance inflation factors gave no indication of multicollinearity in this model.

Serguei Saavedra[1] & Daniel B. Stouffer[2]
[1]Integrative Ecology Group, Estación Biológica de Doñana (EBD-CSIC), Calle Américo Vespucio s/n, E-41092 Sevilla, Spain.
[2]School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand.
email: daniel.stouffer@canterbury.ac.nz

Received 10 October 2012; accepted 4 June 2013.

1. Bastolla, U. *et al.* The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* **458,** 1018–1020 (2009).
2. James, A., Pitchford, J. W. & Plank, M. J. Disentangling nestedness from models of ecological complexity. *Nature* **487,** 227–230 (2012).
3. Thébault, E. & Fontaine, C. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* **329,** 853–856 (2010).
4. Jordano, P., Bascompte, J. & Olesen, J. M. Invariant properties in coevolutionary networks of plant-animal interactions. *Ecol. Lett.* **6,** 69–81 (2003).
5. Fortuna, M. A. *et al.* Nestedness versus modularity in ecological networks: two sides of the same coin? *J. Anim. Ecol.* **79,** 811–817 (2010).
6. Saavedra, S., Stouffer, D. B., Uzzi, B. & Bascompte, J. Strong contributors to network persistence are the most vulnerable to extinction. *Nature* **478,** 233–235 (2011).
7. Stouffer, D. B., Camacho, J., Guimerà, R., Ng, C. A. & Amaral, L. A. N. Quantitative patterns in the structure of model and empirical food webs. *Ecology* **86,** 1301–1311 (2005).
8. Rezende, E. L., Lavabre, J. E., Guimarães, P. R., Jordano, P. & Bascompte, J. Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature* **448,** 925–928 (2007).
9. Almeida-Neto, M., Guimarães, P., Guimarães, P. R. Jr, Loyola, R. D. & Urlich, W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117,** 1227–1239 (2008).
10. Ulrich, W., Almeida-Neto, M. & Gotelli, N. J. A consumer's guide to nestedness analysis. *Oikos* **118,** 3–17 (2009).
11. Gotelli, N. J. & Ulrich, W. Statistical challenges in null model analysis. *Oikos* **121,** 171–180 (2012).
12. Williams, R. J. Biology, methodology or chance? the degree distributions of bipartite ecological networks. *PLoS ONE* **6,** e17645 (2011).
13. Allesina, S. Ecology: the more the merrier. *Nature* **487,** 175–176 (2012).
14. Bascompte, J. & Jordano, P. The structure of plant-animal mutualistic networks: the architecture of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **38,** 567–593 (2007).
15. Zuur, A., Ieno, E. N., Walker, N., Saveiliev, A. A. & Smith, G. M. *Mixed Effects Models and Extensions in Ecology with R* (Springer, 2009).

Author Contributions S.S. and D.B.S. designed the study, performed the simulations, analysed the data and wrote the manuscript.

Competing Financial Interests Declared none.

doi:10.1038/nature12380

Appendix

We randomized the empirical networks with two null models: the probabilistic and fixed (or swap) algorithms[5]. For our purposes here, the key distinction between the two is that the probabilistic model generates random networks with quantitatively more homogeneous degree distributions than those observed empirically ($W_{ij} > 0$) whereas the degree distribution is strictly conserved in networks generated by the fixed model ($W_{ij} \equiv 0$). The statistical analyses presented here were performed in R version 2.15.3 (http://R-project.org/) using the glmer function in package lme4 version 0.999999-0 (http://lme4.r-forge.R-project.org). Code to perform the network randomizations and dynamic simulations in Matlab (http://www.matlab.com/) and the mixed-effects logistic regressions in R (http://R-project.org/) is available from the Dryad Digital Repository at http://dx.doi.org/10.5061/dryad.p2gq8.

James *et al.* reply

REPLYING TO S. Saavedra & D. B. Stouffer *Nature* **500,** http://dx.doi.org/10.1038/nature12380 (2013)

Saavedra and Stouffer[1] claim that the results of James *et al.*[2] are a consequence of the method used to randomize interaction matrices. We recognize the importance of examining alternative randomization schemes and have repeated our analysis using their methods. However, we find no evidence that 'reconfirms the importance of nestedness in mutualistic communities'[1].

Repeating the analysis of figure 2 in ref. 2, using the swap randomization scheme[1,3], which does not change degree distribution, confirms our finding that the persistence of real networks is not related to their nestedness. Although more of the empirical networks are less nested than their randomized counterparts under this scheme, contrary to the accepted result of ref. 4, there is no useful correlation between nestedness and persistence (Fig. 1). Therefore, the 'small but critical step'[1] of accounting for degree heterogeneity does not produce a positive relationship between nestedness and persistence.

The results in figure 1 of ref. 1 represent relationships between nestedness and persistence among randomizations of individual networks. They do not imply that, given two observed networks, the more nested network is more likely to have the higher persistence as claimed in ref. 5. We have performed the general linear mixed model (GLMM) analysis advocated in ref. 1. This shows that >90% of the variance comes from variance between groups (networks) and <10% comes from variance within groups. This highlights the lack of consistency across the groups, and that any effect of nestedness is dwarfed by the random effects of the GLMM.

The NODF[6] definition of nestedness used in refs 1 and 2 is one of several possible metrics. For example, the nestedness metric used in

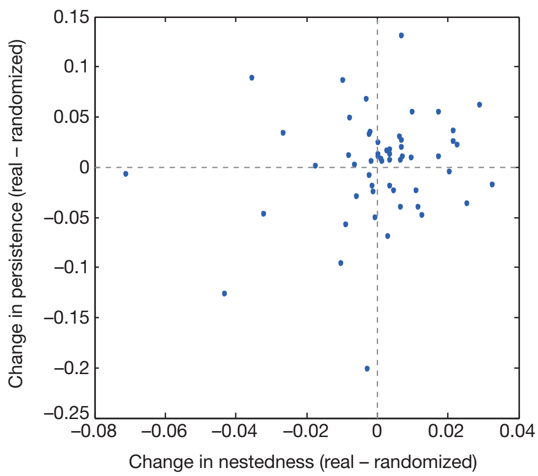


Figure 1 | Accounting for degree distribution does not give a meaningful relationship between nestedness and persistence. Repeating the results of figure 2 in ref. 2 using the swap randomization scheme, which does not change connectance or degree distribution, there is no useful correlation between the change in nestedness (relative to the empirical network) and the change in the persistence of the dynamic model. Each point represents the average of 100 randomizations of an empirical network.

ref. 5 is invariant under the swap randomization scheme of ref. 1, precluding a GLMM approach. Under the metric of ref. 7, the GLMM reveals a negative relationship between nestedness and persistence. That the conclusions of ref. 1 are sensitive to the choice of metric indicates that they cannot be used to draw general conclusions about the effects of nestedness. In contrast, the results in ref. 2 are robust to the choice of metric.

Is nestedness important for predicting persistence in these models? Our results, confirmed by the methods of ref. 1, show that it is less important than: network size; connectance; degree distribution; intrinsic growth rates; competition coefficients; and the strength of the mutualistic interactions. If two ecosystems can be found that share all these properties then, under the specific dynamic model tested here, the more nested ecosystem may (depending how nestedness is defined) be more likely to have a higher persistence. However, if any of these properties differ between the two ecosystems, then any effect of nestedness is likely to be unimportant.

In conclusion, nestedness is an interesting abstract network property that undoubtedly influences the statistical behaviour of large systems of differential equations[5]. However, general conclusions allowing nestedness to be used as a predictor of empirical biodiversity cannot currently be justified.

Alex James[1], Jonathan W. Pitchford[2] & Michael J. Plank[1]
[1]Biomathematics Research Centre, University of Canterbury, Private Bag 4800, Christchurch 8040, New Zealand.
email: alex.james@canterbury.ac.nz
[2]York Centre for Complex Systems Analysis, and Departments of Biology and Mathematics, University of York, Heslington, York YO10 5DD, UK.

1. Saavedra, S. & Stouffer, D. B. ''Disentangling nestedness'' disentangled. *Nature* **500,** http://dx.doi.org/nature12380 (2013).
2. James, A., Pitchford, J. W. & Plank, M. J. Disentangling nestedness from models of ecological complexity. *Nature* **487,** 227–230 (2012).
3. Fortuna, M. A. *et al.* Nestedness versus modularity in ecological networks: two sides of the same coin? *J. Anim. Ecol.* **79,** 811–817 (2010).
4. Bascompte, J., Jordano, P., Melián, C. J. & Olesen, J. The nested assembly of plant–animal mutualistic networks. *Proc. Natl Acad. Sci. USA* **100,** 9383–9387 (2003).
5. Bastolla, U. *et al.* The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* **458,** 1018–1020 (2009).
6. Almeida-Neto, M., Guimarães, P., Guimarães, P. R., Loyola, R. D. & Ulrich, W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **117,** 1227–1239 (2008).
7. Staniczenko, P. P. A., Kopp, J. C. & Allesina, S. The ghost of nestedness in ecological networks. *Nature Commun.* **4,** 1391 http://dx.doi.org/10.1038/ncomms2422 (2013).

doi:10.1038/nature12381